

**Entwicklung eines
abstimmbaren Hochleistungslasersystems
für die nichtlineare Spektroskopie
im fernen Infrarot**

DISSERTATION

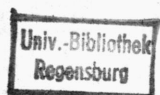
**ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER
NATURWISSENSCHAFTLICHEN FAKULTÄT II - PHYSIK
DER UNIVERSITÄT REGENSBURG**

**VORGELEGT VON
ULRICH WERLING
AUS REGENSBURG**

1988

Di 88/45

200/ UH 5710 W489 +2



6477 786

Promotionsgesuch eingereicht am : 2.2.1988
Die Arbeit wurde angeleitet von : Prof. Dr. K.F. Renk
Prüfungsausschuß : Prof. Dr. J. Keller , Vorsitzender
Prof. Dr. K.F. Renk , Erstgutachter
Prof. Dr. A. Penzkofer , Zweitgutachter
Prof. Dr. H. Hoffmann

Inhaltsverzeichnis

Einleitung	1
I. Aufbau und Untersuchung eines Hochleistungslasersystems zur abstimmbaren Erzeugung von Ferninfrarotstrahlung mit Hilfe stimulierter Ramanemission in CH_3F	4
I.1 Der Hochdruck- CO_2 -Laser	4
1.1 Kleinsignalverstärkung eines Hochdruck- CO_2 -Lasers	5
1.1.1 Energieniveauschema eines CO_2 -Moleküls	5
1.1.2 Spektrale Breite der Laserübergänge	8
1.1.3 Temperaturmodell der Kleinsignalverstärkung	10
1.2 Anregung eines Hochdruck- CO_2 -Lasers durch elektrische Gasent- ladungen	14
1.2.1 Prinzip der elektrischen Gasentladung	15
1.2.2 Bedingungen für die effiziente Anregung des Lasergases in einer elektrischen Gasentladung	16
1.2.3 Vorionisierung bei selbständigen Gasentladungen	20
1.3 Aufbau des Hochdruck- CO_2 -Lasers	22
1.3.1 Elektroden	23
1.3.2 Vorionisierung	26
1.3.3 Hochspannungsversorgung	29
1.3.4 Mechanischer Aufbau	32
1.3.5 Resonatorkonfiguration	35
1.4 Eigenschaften des Hochdruck- CO_2 -Lasers	37
1.4.1 Abstimmcharakteristik	38
1.4.2 Zeitliche Struktur der Laserpulse	42
1.4.3 Breitbandemission	45
1.5 Zusammenstellung der Eigenschaften des Hochdruck- CO_2 -Lasers und Ausbaumöglichkeiten	47

I.2	Der CH_3F -Raman-Laser	50
2.1	Stimulierte Ramanemission in CH_3F	51
2.1.1	Energieniveauschema eines CH_3F -Moleküls	51
2.1.2	Kleinsignalverstärkung für die stimulierte Ramanstreuung in CH_3F	55
2.2	Aufbau des CH_3F -Raman-Lasers	63
2.3	Eigenschaften des CH_3F -Raman-Lasers	66
2.3.1	Emissionscharakteristik	67
2.3.2	Wirkungsgrad	71
2.4	Übersicht über die Eigenschaften des CH_3F -Raman-Lasers	75
I.3	Experimentelle Anordnung und Techniken für die FIR-Transmissions- spektroskopie mit Hilfe des CH_3F -Raman-Lasers	77
3.1	Optischer Aufbau	77
3.2	Meßwerterfassungssystem	80
3.3	Experimentelle Methoden zur Bestimmung des Absolutwertes der Transmission	84
II.	Nichtlineare Absorption von Ferninfrarotstrahlung durch eine resonante Band- mode in Sauerstoff-dotierten Siliziumkristallen	86
II.1	Interstitieller Sauerstoff in Siliziumkristallen	86
1.1	Bedeutung von Sauerstoff als Verunreinigung in Silizium bei der Produktion von Halbleiterbauelementen	86
1.2	Einbaulage von Sauerstoff in Siliziumkristallen	87
1.3	Infrarotabsorptionseigenschaften von Sauerstoff-dotiertem Silizium	88
1.4	Resonante Streuung akustischer Phononen an interstitiellen Sauerstoff- atomen in Silizium	93
II.2	Transmissionsspektrum von Sauerstoff-dotiertem Silizium im fernen Infrarot	98
II.3	Nichtlineare Absorption von Ferninfrarotstrahlung in Sauerstoff-dotiertem Silizium	103
3.1	Experimentelle Ergebnisse	103

3.2	Analyse der experimentellen Ergebnisse	107
3.2.1	Modellrechnung für die Abhängigkeit der Transmission eines Sauerstoff-dotierten Siliziumkristalls von der FIR-Pulsenergie für den Fall vernachlässigbarer Dämpfung der resonanten Bandmode	109
3.2.2	Mögliche Dämpfungsmechanismen für die resonante Bandmode . . .	116
3.2.3	Laufzeitexperimente zur Untersuchung der Ausbreitungseigen- schaften der durch die FIR-Absorption erzeugten Phononen: Hinweis auf eine räumlich-spektrale Diffusion	119
3.2.4	Erweiterung der Modellrechnung mit spektraler Diffusion der resonanten Phononen als Dämpfungsmechanismus der resonanten Bandmode	125
3.3	Ausblick	129
	Zusammenfassung	132
	Literaturverzeichnis	136

Teilergebnisse dieser Arbeit wurden veröffentlicht in:

Wan Chong-Yi, U. Werling und K.F. Renk

Study of broadband emission of an ultraviolet preionized 20-atm CO₂ laser

J. Appl. Phys. **57**, 990 (1985).

U. Werling, Wan Chong-Yi und K.F. Renk

Tuning characteristics of a UV-preionized 20-atmospheres CO₂ laser

Intern. J. Infrared and Millimeter Waves **6**, 449 (1985).

U. Werling, K.F. Renk und Wan Chong-Yi

Tuning characteristics of a high pressure CO₂ laser pumped CH₃F Raman laser

Intern. J. Infrared and Millimeter Waves **7**, 881 (1986).

U. Werling, Wan Chong-Yi und K.F. Renk

Efficiency of a high pressure CO₂ laser pumped CH₃F Raman laser

IEEE J. Quantum Electron. **QE-22**, 1024 (1986).

U. Werling und K.F. Renk

Propagation of high-frequency phonons in oxygen doped silicon

Proc. 18th Intern. Conf. on Low Temperature Physics, Kyoto 1987: Japanese J. of Appl. Phys. **26**, 671 (1987).

Einleitung

Die nichtlineare Spektroskopie bietet die Möglichkeit, Informationen über mikroskopische Materialeigenschaften zu erhalten, die durch klassische optische Methoden nicht zugänglich sind. So können beispielsweise mit Hilfe der Zweiphotonenspektroskopie Übergänge zwischen Energiezuständen, die für elektrische Dipolstrahlung verboten sind, untersucht werden; die sättigbare Absorption ermöglicht die Bestimmung der Energierelaxationszeit T_1 angeregter Zustände, spektrales Lochbrennen liefert die homogene Linienbreite und damit die Phasenrelaxationszeit T_2 in inhomogen verbreiterten Systemen. Kohärente transiente Effekte wie freier Induktionszerfall und Photonenecho gestatten die Untersuchung der Dynamik angeregter Zustände wie z.B. Relaxations- und Energietransferprozesse.

Während sehr viele Experimente und Techniken der nichtlinearen Spektroskopie im sichtbaren Spektralbereich, im nahen und im mittleren Infrarot veröffentlicht wurden - einige Beispiele findet man in /1,2/, ist ihre Zahl im ferninfraroten (FIR) Spektralbereich (Wellenlängenbereich $100\ \mu\text{m}$ - $1\ \text{mm}$ bzw. Frequenzbereich von $10\ \text{cm}^{-1}$ - $100\ \text{cm}^{-1}$) noch sehr gering. Einige der prominentesten Arbeiten sind die Beobachtung der Zweiphotonenzyklotronresonanz und des $1s \rightarrow 2s$ Zweiphotonenübergangs an flachen Donatoren in GaAs /3/, die Sättigung von Zyklotronresonanz und Übergängen zwischen Zuständen flacher Donatoren in GaAs /4,5/ und die nichtlineare Absorption von FIR-Strahlung durch die Zyklotronresonanz in einem zweidimensionalen Elektronengas in GaAs/GaAl_xAs_{1-x} Heterostrukturen /6/.

Die Ursache der spärlichen Zahl von Experimenten zur nichtlinearen Spektroskopie im fernem Infrarot ist das Fehlen geeigneter Hochleistungslaser in diesem Spektralbereich. Obwohl seit der ersten Demonstration eines optisch gepumpten FIR-Lasers im Jahre 1970 durch Chang /7/ weit über 1000 diskrete FIR-Laserlinien im Bereich von $100\ \mu\text{m}$ bis $1\ \text{mm}$ gefunden wurden /8/, ist die Untersuchungsmöglichkeit eines physikalischen Problems bei einer bestimmten FIR-Frequenz immer noch an die zufällige Übereinstimmung dieser Frequenz mit einer der diskreten FIR-Laserlinien gebunden. Selbst wenn ein solcher Fall gegeben ist, erfordert eine Frequenzvariation meist den Austausch des FIR-Lasermediums, der Anregungsmethode und der Laserkonstruktion. Eine in einem weiten Bereich abstimmbare kohärente FIR-Strahlungsquelle würde daher neue Möglichkeiten für die nichtlineare Spektroskopie im ferninfraroten Spektralbereich eröffnen.

In den letzten Jahren wurden mehrere Methoden zur abstimmbaren Erzeugung von FIR-Strahlung entwickelt: Der Freie-Elektronen-Laser (FEL) benutzt einen gepulsten Strahl von relativistischen Elektronen in einem transversalen, räumlich periodischen Magnetfeld zur Verstärkung von kollinear verlaufender kohärenter Strahlung. Seine Emissionsfrequenz hängt

von der Periode des Magnetfelds und der Elektronenenergie ab und kann durch Variation der letzteren geändert werden. Ein Beispiel für einen im ferninfraroten Spektralbereich arbeitenden FEL ist der mit einem Van de Graaff-Beschleuniger ausgerüstete UCSB-FEL (University of California at Santa Barbara). Er kann FIR-Strahlung im Frequenzbereich von 20 cm^{-1} bis 80 cm^{-1} mit einer Pulsleistung von $0.1 - 1\text{ kW}$ (Pulslänge $0.5 - 4\text{ }\mu\text{sec}$) erzeugen [5,9]. Die FIR-Emission durch heiße Ladungsträger in Halbleitern bietet eine weitere Möglichkeit, abstimmbare FIR-Strahlung zu erzeugen. Unter bestimmten Bedingungen kann beispielsweise in hochreinem p-Germanium in gekreuzten elektrischen und magnetischen Feldern eine Überbesetzung des leichten Löcherbandes gegenüber dem schweren Löcherband erreicht werden. Eine breitbandige ($\sim 30\text{ cm}^{-1}$) stimulierte FIR-Emission, deren Zentrum von den Werten des gepulsten elektrischen und statischen magnetischen Feldes bestimmt wird, wurde von Andronov et al. [10,11] im Wellenlängenbereich von $75\text{ }\mu\text{m}$ bis $250\text{ }\mu\text{m}$ mit Pulsleistungen von 10 W beobachtet. Durch Frequenzmischung der im sichtbaren Spektralbereich liegenden Strahlung von 2 Dylasern in LiNbO_3 wurde von Berg et al. [12] FIR-Strahlung abstimmbare im Bereich von 20 cm^{-1} bis 200 cm^{-1} erzeugt. Da einer der beiden Dylaser Pikosekundenpulse emittierte, lagen auch die Pulsdauern der FIR-Strahlung im Pikosekundenbereich. Spitzenleistungen von $0.3 - 1\text{ kW}$ wurden erreicht; die Bandbreite der FIR-Strahlung betrug $\sim 10\text{ cm}^{-1}$ entsprechend der zeitlichen Dauer des Pulses.

Die eben geschilderten Methoden erlauben zwar alle eine abstimmbare Erzeugung von FIR-Strahlung, erfordern aber entweder einen großen experimentellen Aufwand wie im Falle eines FEL's oder sind infolge ihrer charakteristischen Eigenschaften wie Pulsleistung oder spektrale Breite nur bedingt für die nichtlineare Spektroskopie geeignet. Es gibt jedoch noch eine weitere, relativ einfache Möglichkeit, kohärente FIR-Strahlung hoher Pulsleistung zu erzeugen, nämlich die stimulierte Ramanstreuung an molekularen Gasen. Verwendet man dafür eine abstimmbare Pumpquelle, läßt sich deren Abstimmbarkeit in den ferninfraroten Spektralbereich übertragen. Die erste Demonstration dieses Verfahrens erfolgte von Frey et al. [13,14] an gasförmigem HCl bzw. HF. Sie benutzten dabei eine relativ aufwendige Pumpquelle, nämlich die zweifach in H_2 -Gas Raman-verschobene Strahlung eines von einem Rubinlaser gepumpten abstimmbaren Dylasers und erreichten im Falle von HF-Gas eine abstimmbare Erzeugung von FIR-Strahlung im Frequenzbereich von 40 cm^{-1} bis 200 cm^{-1} in einzelnen Intervallen von 5 cm^{-1} Breite. Ein experimentell einfacheres Verfahren wurde möglich, als in den letzten Jahren kontinuierlich abstimmbare Hochdruck- CO_2 -Laser mit hoher Ausgangsleistung entwickelt wurden. Derartige Laser, deren Abstimmbarkeit eine Folge des Überlapps der druckverbreiterten Rotations-Schwingungs-Übergänge der CO_2 -Moleküle ist, sind eine sehr gut geeignete Pumpquelle zur abstimmbaren Erzeugung von FIR-Strahlung durch stimulierte Ramanstreuung. So fällt beispielsweise der $9R$ -Emissionsbereich eines solchen Lasers mit dem ν_3 -R-Absorptionszweig in gasförmigem Methylfluorid (CH_3F) zusammen. Theoretische Rechnungen von Biron et al. [15] lassen in diesem Fall eine sehr

hohe Kleinsignalverstärkung für die stimulierte Ramanstreuung erwarten. Tatsächlich gelang es mehreren Gruppen (Mathieu et al. /16/, Danly et al. /17/) in ersten Experimenten, FIR-Strahlung im Frequenzbereich von 25 cm^{-1} bis 45 cm^{-1} mit einer maximalen Pulsleistung von 4 kW zu erzeugen.

Auch die vorliegende Arbeit beschäftigt sich mit der abstimmbaren Erzeugung von FIR-Strahlung durch stimulierte Ramanstreuung in CH_3F . Mit Hilfe eines Hochdruck- CO_2 -Lasers, der mit einigen technischen Neuerungen versehen war, konnte FIR-Strahlung im Bereich von 23 cm^{-1} bis 45 cm^{-1} mit maximalen Pulsleistungen von über 7 kW erzeugt werden. Gleichzeitig wird erstmals die Anwendungsmöglichkeit dieser Strahlung für die Untersuchung nichtlinearer Phänomene in Festkörpern im fernen Infrarot, und damit eine erste Nutzung in der Festkörperspektroskopie, demonstriert.

Die Arbeit gliedert sich wie folgt: Im ersten Abschnitt wird zunächst der Aufbau und die Funktionsweise des durch eine elektrische Gasentladung mit UV-Vorionisierung angeregten und bei einem Druck von 20 atm betriebenen Hochdruck- CO_2 -Lasers beschrieben. Insbesondere wird eine neuartige Anordnung für die UV-Vorionisierung vorgestellt, die eine besonders einfache und kompakte Laserkonstruktion ermöglichte. Die charakteristischen Eigenschaften des Lasers werden diskutiert und mit denen anderer bekannter Konstruktionen verglichen. Der zweite Teil dieses Abschnitts beschäftigt sich mit den Eigenschaften eines von dem Hochdruck- CO_2 -Laser gepumpten CH_3F -Raman-Lasers. Erstmals wird eine vollständige Emissionscharakteristik eines solchen Lasers vorgestellt. Ihre wesentlichen Eigenschaften werden besprochen, Untersuchungen des Wirkungsgrades für die Umsetzung von CO_2 -Laserstrahlung in FIR-Strahlung werden durchgeführt. Der zweite Abschnitt der Arbeit berichtet von einem Einsatz des CH_3F -Raman-Lasers in der nichtlinearen Spektroskopie. Es wurde erstmals ein nichtlineares Verhalten der Absorption von FIR-Strahlung infolge der Anregung einer Schwingungsmode im Frequenzbereich der hochfrequenten akustischen Phononen eines Festkörpers bei hohen FIR-Pulsleistungen beobachtet. Dabei wurde speziell die durch interstitiell in Silizium eingebaute Sauerstoffatome verursachte resonante Bandmode in Silizium untersucht, deren Resonanzfrequenz mit bisherigen FIR-Lasern nicht zugänglich war. Der Vergleich der experimentellen Ergebnisse mit Modellrechnungen erlaubte Aussagen über mögliche Dämpfungsmechanismen der resonanten Bandmode.

I. Aufbau und Untersuchung eines Hochleistungslasersystems zur abstimmbaren Erzeugung von Ferninfrarotstrahlung mit Hilfe stimulierter Ramanemission in CH_3F

In diesem Abschnitt werden Aufbau und Funktionsweise des im Laufe der vorliegenden Arbeit entwickelten, aus einem Hochdruck- CO_2 -Laser und CH_3F -Raman-Laser bestehenden Hochleistungslasersystems zur abstimmbaren Erzeugung von FIR-Strahlung beschrieben. Die charakteristischen Eigenschaften des Lasersystems werden vorgestellt und diskutiert.

I.1. Der Hochdruck- CO_2 -Laser

1971 wurde von Bagratashvili et al. /18/ erstmals vorgeschlagen, das Emissionsspektrum von CO_2 -Lasern, das aus einer Vielzahl von diskreten Linien im Wellenlängenbereich um $9\text{ }\mu\text{m}$ bzw. $10\text{ }\mu\text{m}$ besteht, durch den Betrieb bei hohen Gasdrücken zu vergrößern. Die Verbreiterung der Emissionslinien bei genügend hohen Drücken führt zu einem Überlapp dieser Linien und ermöglicht somit einen Betrieb des Lasers auch bei Frequenzen, die zwischen den Liniencentren liegen. Kontinuierliche Erzeugung kohärenter Infrarotstrahlung im obigen Wellenlängenbereich ist möglich und wurde in den letzten Jahren auch von vielen Arbeitsgruppen demonstriert. Verschiedenste Anregungstechniken, wie z.B. optisches Pumpen /19,20/, von einem Elektronenstrahl kontrollierte nichtselbständige elektrische Gasentladungen /24-24/ bzw. durch UV-Strahlung vorionisierte selbständige Gasentladungen /16,17,25-29/ wurden erprobt. In dieser Arbeit wird ein mit dem letztgenannten Verfahren angeregter, bei einem Druck von 20 atm arbeitender Hochdruck- CO_2 -Laser präsentiert, bei dem eine neue Anordnung für die UV-Vorionisierung entwickelt wurde, die eine besonders kompakte Laserkonstruktion ermöglichte.

Es sei noch bemerkt, daß Hochdruck- CO_2 -Laser neben der Möglichkeit der Erzeugung abstimmbarer FIR-Strahlung mittels stimulierter Ramanemission, von der in dieser Arbeit berichtet wird, eine Vielzahl weiterer Anwendungsmöglichkeiten besitzen. Genannt sei als Beispiel das effiziente resonante Pumpen molekularer Gase zur Erzeugung kohärenter Strahlung im mittleren und fernen Infrarot /30,31/, die Erzeugung und Verstärkung von ultrakurzen Laserpulsen (Pulsdauer $< 100\text{ psec}$) im Wellenlängenbereich um $9\text{ }\mu\text{m}$ bzw. $10\text{ }\mu\text{m}$ /32,33/ und die Erzeugung breitbandiger Strahlung im genannten Wellenlängenbereich, die sich be-

sonders für Intracavity-Spektroskopie /34/ oder auch für Bestimmung von gasförmigen Verunreinigungen in der Erdatmosphäre (LIDAR) /35/ verwenden läßt.

Bevor nun der Aufbau und die charakteristischen Eigenschaften des in dieser Arbeit entwickelten Hochdruck-CO₂-Lasers näher besprochen werden, seien kurz die spektroskopischen und molekularkinetischen Grundlagen, die die Kleinsignalverstärkung eines unter hohem Druck betriebenen CO₂-Lasers bestimmen, diskutiert.

1.1 Kleinsignalverstärkung eines Hochdruck-CO₂-Lasers

1.1.1 Energieniveauschema eines CO₂-Moleküls

Das Emissionsspektrum eines CO₂-Lasers liegt im Frequenzbereich von 900 cm⁻¹ bis 1100 cm⁻¹ und wird bestimmt durch Übergänge zwischen Rotations-Schwingungs-Zuständen eines CO₂-Moleküls im elektronischen Grundzustand. Als lineares dreiatomiges Molekül besitzt ein CO₂-Molekül bekanntlich 3 verschiedene Normalschwingungen, nämlich die symmetrische Streckschwingung ν_1 , die zweifach entartete Biegeschwingung ν_2 und die asymmetrische Streckschwingung ν_3 . Die Schwingungszustände eines CO₂-Moleküls lassen sich somit durch ein Tripel $|v_1, v_2^l, v_3\rangle$ charakterisieren, wobei v_1, v_2 bzw. v_3 die Schwingungsquantenzahlen der Normalschwingungen ν_1, ν_2 bzw. ν_3 sind. Da die Biegeschwingung infolge der zwei möglichen zueinander senkrechten Auslenkungsrichtungen einen zweidimensionalen Oszillator darstellt, ist zur Charakterisierung ihrer Energiezustände neben der Schwingungsquantenzahl v_2 eine weitere Quantenzahl l ($l = v_2, v_2 - 2, \dots, -v_2$) nötig, die den Winkelanteil der Eigenfunktionen beschreibt /36/. Die Energien der einzelnen Zustände sind gegeben durch

$$E_S(v_1, v_2^l, v_3) = h \left(\sum_{i=1}^3 \nu_i \left(v_i + \frac{d_i}{2} \right) + \sum_{\substack{i,k=1 \\ k \geq i}}^3 x_{ik} \left(v_i + \frac{d_i}{2} \right) \left(v_k + \frac{d_k}{2} \right) + gl^2 \right) \quad (1)$$

wobei d_i den Entartungsgrad der Normalschwingung ν_i angibt. Die Größen x_{ik} beschreiben die Wechselwirkung zwischen den Normalschwingungen aufgrund anharmonischer Anteile im Potential, die Größe g gibt eine Energiekorrektur für die Schwingungszustände v_2 der Normalmode ν_2 für verschiedene Quantenzahlen l an, die ebenfalls eine Folge der Anhar-

nizität des Potentials ist. Für die Werte dieser Größen gilt /36/:

$$\begin{aligned} \nu_1 &= 1351.2 \text{ cm}^{-1} & \nu_2 &= 672.2 \text{ cm}^{-1} & \nu_3 &= 2396.4 \text{ cm}^{-1} \\ x_{11} &= -0.3 \text{ cm}^{-1} & x_{22} &= -1.3 \text{ cm}^{-1} & x_{33} &= -12.5 \text{ cm}^{-1} \\ x_{12} &= 5.7 \text{ cm}^{-1} & x_{13} &= -21.9 \text{ cm}^{-1} & x_{23} &= -11.0 \text{ cm}^{-1} \\ g &= 1.7 \text{ cm}^{-1} \end{aligned}$$

Abb. 1 zeigt den für das Verständnis des Emissionsspektrums eines CO₂-Lasers nötigen Teil des Schwingungsniveauschemas. Die Laserübergänge erfolgen vom Zustand $|00^1_1\rangle$ in die mit $|10^0\rangle|02^0\rangle|I,II\rangle$ bezeichneten Niveaus. Diese gehen aus den Zuständen $|10^0\rangle$ und $|02^0\rangle$ hervor, die nach Gl.(1) nahezu gleiche Energie haben, d.h. zufällig entartet sind. Eine Wechselwirkung zwischen beiden Zuständen wiederum aufgrund anharmonischer Anteile im Potential führt zu einer Abstoßung der beiden Energiezustände; man sagt, die Zustände sind in Fermiresonanz /36/. Es bilden sich neue Eigenzustände, die Wellenfunktionen der Form $a \cdot |10^0\rangle \pm b \cdot |02^0\rangle$ besitzen und mit $|10^0\rangle|02^0\rangle|I,II\rangle$ bezeichnet werden. Ihre Energien liegen bei $E_I = 1388.3 \text{ cm}^{-1}$ bzw. $E_{II} = 1285.5 \text{ cm}^{-1}$, gehorchen also nicht mehr Gl.(1). (Die Zustände $|10^0\rangle$ und $|02^0\rangle$ befinden sich nicht in Fermiresonanz, da diese nur zwischen Zuständen gleichen Symmetrietyps auftreten kann /36/). Der Einfachheit halber seien die besagten Niveaus dennoch mit $|10^0\rangle$ und $|02^0\rangle$ bezeichnet.

Infolge der Rotationsmöglichkeit eines CO₂-Moleküls um eine Achse senkrecht zur Molekülachse ist mit jedem Schwingungsniveau eine Unterstruktur von Rotationsniveaus, die durch die Rotationsquantenzahl J gekennzeichnet werden, verbunden. Symmetriebetrachtungen über den Austausch der Sauerstoffkerne innerhalb des CO₂-Moleküls (das Isotop ¹²C¹⁶O₂ sei vorausgesetzt!) zeigen, daß bei Schwingungszuständen, die symmetrisch bezüglich des Austausches sind (z.B. die Σ_g -Zustände $|10^0\rangle, |02^0\rangle$), nur Rotationszustände mit geradem J auftreten. Bei bezüglich des Austausches antisymmetrischen Schwingungszuständen (z.B. der Σ_u -Zustand $|00^1_1\rangle$) gibt es nur Rotationszustände mit ungeradem J /36/. Die Gesamtenergie eines CO₂-Moleküls setzt sich damit aus der Schwingungsenergie E_S und der Rotationsenergie

$$E_R(J) = B J(J+1) \quad (2)$$

zusammen. B ist die Rotationskonstante; sie hängt vom Trägheitsmoment des Moleküls für die Rotation um eine Achse senkrecht zur Molekülachse ab und ist damit abhängig vom zugehörigen Schwingungszustand. Für die an den Laserübergängen beteiligten Schwingungsniveaus gilt:

$$B_{|00^1_1\rangle} = 0.3866 \text{ cm}^{-1}, \quad B_{|10^0\rangle} = 0.3897 \text{ cm}^{-1}, \quad B_{|02^0\rangle} = 0.3899 \text{ cm}^{-1}$$

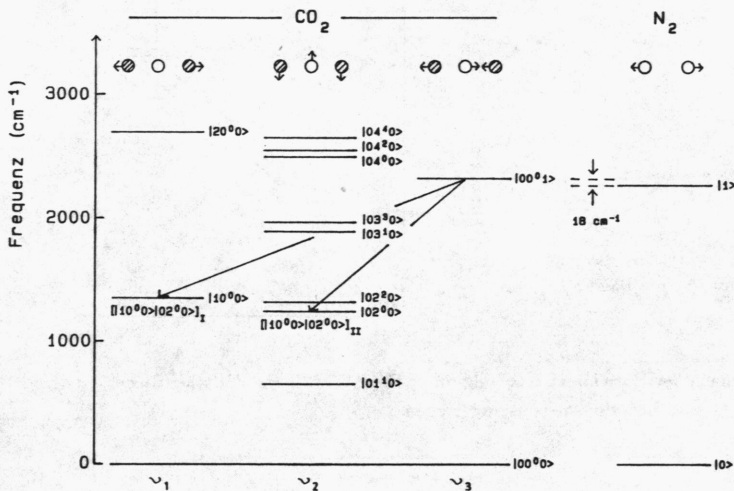
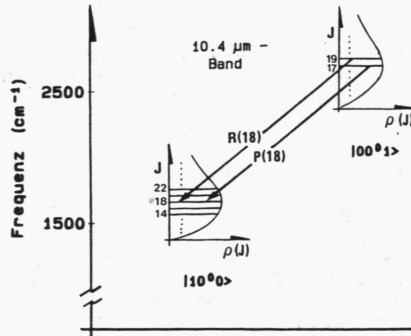


Abb. 1 Schwingungsniveauschema für CO₂- bzw. N₂-Moleküle.

Jeder der beiden Laserübergänge $|00^{\circ}1\rangle \rightarrow |10^{\circ}0\rangle$ bzw. $|00^{\circ}1\rangle \rightarrow |02^{\circ}0\rangle$ besteht somit aus einer Vielzahl einzelner Linien infolge der Unterstruktur der Schwingungsniveaus durch die der Schwingung überlagerten Rotationen. Sie werden nach der Lage ihres Linienschwerpunkts als das 10.4 μm bzw. 9.4 μm Band bezeichnet. Die gültige Auswahlregel für die Änderung der Rotationsquantenzahl ist $\Delta J = \pm 1$ ($\Delta J = 0$ ist nicht möglich, da das obere Laserniveau nur ungerade, die unteren jedoch nur gerade J besitzen). Jedes der Bänder zerfällt somit in 2 Zweige, den höherenergetischen *R*-Zweig, der aus Übergängen mit $\Delta J = +1$ besteht, sowie den niederenergetischen *P*-Zweig, der von den Übergängen $\Delta J = -1$ gebildet wird. Es ist dabei üblich, einen Übergang durch die Quantenzahlen des tieferliegenden Energiezustandes zu charakterisieren; auch die Änderung von J bezieht sich auf diesen Zustand. Beispiele für die Notation anhand von 2 Laserlinien innerhalb des 10*R*- bzw. 10*P*-Zweiges zeigt Abb. 2.

Der Abstand benachbarter Linien innerhalb der *R*-Zweige beträgt nach Gl.(2) etwa 1.3 cm^{-1} (39 GHz), innerhalb der *P*-Zweige etwa 1.8 cm^{-1} (54 GHz). Eine genaue Tabellierung der Laserlinien findet man in [37/. In Abb. 2 ist auch die relative Besetzung $\rho(J) = \frac{N_J}{N}$ der einzelnen Rotationsniveaus bezüglich der Besetzung N des zugehörigen Schwingungs-

Abb. 2
Rotations-Schwingungs-Übergänge
innerhalb des 10.4 μm Bandes.



zustandes nach der Boltzmannverteilung unter Berücksichtigung des statistischen Gewichts $g_J = 2J + 1$ eines Rotationsniveaus angetragen

$$\rho(J) = \frac{N_J}{N} \sim (2J + 1) \cdot \exp\left(-\frac{E_R(J)}{kT}\right) \quad (3)$$

k ist die Boltzmannkonstante, T die Temperatur. Bei $T = 300$ K liegt das Maximum der Besetzungsverteilung etwa bei $J = 16$. Maximale Kleinsignalverstärkung ist damit für die Laserübergänge $R(16)$ bzw. $P(16)$ zu erwarten.

1.1.2 Spektrale Breite der Laserübergänge

Die Lebensdauer des angeregten Schwingungszustandes $|00^01\rangle$ gegenüber der spontanen Emission von Photonen beträgt im Falle des Übergangs $|00^01\rangle \rightarrow |00^00\rangle$ etwa 0.1 sec, für alle anderen Übergänge liegt er zwischen 0.5 sec und 5 sec /38/. Die Linienverbreiterung aufgrund der natürlichen Lebensdauer kann damit gegenüber anderen Verbreiterungsmechanismen wie Doppler- bzw. Stoßverbreiterung vernachlässigt werden. Die Dopplerverbreiterung beträgt bei 400 K etwa 150 MHz /35/ und ist der dominante Verbreiterungsmechanismus bei in Dauerstrich betriebenen Niederdruck- CO_2 -Lasern, deren Betriebsdruck unter 30 Torr liegt. Bei Betriebsdrücken über 50 Torr wird die Linienbreite der Übergänge nur durch die Stoß- bzw. Druckverbreiterung bestimmt. Stoßprozesse unter den Gasmolekülen zerstören eine feste Phasenbeziehung in den Wellenfunktionen der einzelnen Moleküle innerhalb einer charakteristischen Zeit T_2 (sog. Phasenrelaxationszeit) und führen dadurch zu einer Verbreiterung der Emissionslinien. Es läßt sich zeigen, daß die Linienform im Falle der

Stoßverbreiterung durch eine Lorentzfunktion gegeben ist, für deren Halbwertsbreite $\Delta\nu$ gilt /39/:

$$\Delta\nu = \frac{1}{\pi T_2} \quad (4)$$

Die Druckverbreiterung ist für beide Emissionsbänder eines CO₂-Lasers nahezu gleich und unabhängig von der Rotationsquantenzahl J /40,41/. Für ein bei CO₂-Lasern übliches Lasergasgemisch aus He, N₂ und CO₂ (auf die Bedeutung von He und N₂ wird noch eingegangen!) ist die Druckverbreiterung nach Abrams /42/, der eine kritische Analyse verschiedener experimenteller Werte durchführte, gegeben durch

$$\Delta\nu = 7.58 \cdot (\psi_{CO_2} + 0.73\psi_{N_2} + 0.64\psi_{He}) \cdot p \cdot \sqrt{\frac{300}{T}} \quad (5)$$

Hierbei ist $\Delta\nu$ die Linienbreite in MHz (volle Breite bei halber Höhe [FWHM]), p der Gesamtdruck der Gasmischung in Torr und T die Temperatur in K; ψ_{CO_2} , ψ_{N_2} bzw. ψ_{He} geben den Bruchteil von CO₂, N₂ und He innerhalb der Gasmischung an. Einige Beispiele für verschiedene Gasmischungen bei verschiedenen Drücken sind in Tab. 1 angeführt ($T = 300$ K).

Gasmischung CO ₂ :N ₂ :He	Druck	Linienbreite
1:1:8	1 atm	3.9 GHz
3:1:60	10 atm	37 GHz
3:1:60	20 atm	75 GHz
3:1:60	25 atm	94 GHz

Tab. 1 Spektrale Breite der Rotations-Schwingungs-Übergänge für verschiedene Drücke und Gasmischungen.

Der erste Eintrag in der Tabelle kennzeichnet Werte, die bei gepulsten TEA-CO₂-Lasern (Transversely Excited Atmospheric Pressure Laser) üblich sind. Hier sind die Linienbreiten noch sehr viel kleiner als der Abstand benachbarter Linien in den *R*- bzw. *P*-Zweigen. Das Emissionsspektrum eines TEA-CO₂-Lasers besteht somit noch immer aus diskreten Linien. Bei 10 atm erreicht die Halbwertsbreite den Abstand benachbarter Linien in den *R*-Zweigen, jedoch noch nicht in den *P*-Zweigen. (Die angegebene Gasmischung wurde in dem in dieser Arbeit vorgestellten Hochdruck-CO₂-Laser verwendet). Bei Drücken über 20 atm tritt voller Überlapp benachbarter Linien in allen Zweigen ein. Ein kontinuierliches Emissionsspektrum ist die Folge. Es sei noch bemerkt, daß eine Berechnung des Absorptionskoeffizienten bei einer bestimmten Frequenz ν durch einfache Summation über durch ein Lorentzprofil beschriebene Absorptionslinien geringere Werte liefert als bei hohen Drücken

beobachtet wurde /43/. Nach Miller /43/ liegt die Ursache in einer Wechselwirkung zwischen benachbarten Emissionslinien infolge von Stößen, die zu von der Lorentzform abweichenden Linienformen führt und eine etwa 15% stärkere Absorption bei 20 atm zur Folge hat, als bei einer Summation über Lorentzprofile zu erwarten wäre.

1.1.3 Temperaturmodell der Kleinsignalverstärkung

Da die kinetische Translationsenergie $\frac{3}{2} k T$ der Gasmoleküle bei $T = 300$ K etwa 300 cm^{-1} beträgt, also sehr groß gegen den Abstand benachbarter Rotationszustände (Gl.(2)) ist, erfolgt durch Stöße der CO_2 -Moleküle untereinander bzw. mit den restlichen in der Gas-mischung enthaltenen Teilchen ein sehr schneller Austausch ihrer Rotationsenergie mit der Translationsenergie der Gasmoleküle (R-T-Transfer). Nach Gl.(4) und den Werten für die Druckverbreiterung in Tab. 1 findet der Austausch bei einem Gasdruck von 20 atm in Zeiten von weniger als 10 psec statt. Die Temperatur innerhalb des Rotationssystems kann somit stets gleich der kinetischen Gastemperatur T angesetzt werden.

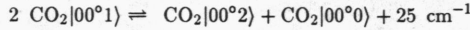
Der Austausch von Vibrationsenergie der CO_2 -Moleküle mit der Translationsenergie der Gasmoleküle (V-T-Transfer) erfolgt wesentlich langsamer, da die mittlere kinetische Energie der Gasmoleküle kleiner als der Abstand der Schwingungszustände (Gl.(1)) ist. Eine Zusammenstellung der Ratenkonstanten für den V-T-Transfer für die ν_2 - bzw. ν_3 -Normalschwingung durch Stöße der CO_2 -Moleküle untereinander bzw. mit N_2 - oder He-Teilchen ist in Tab. 2 gegeben.

Stoßpartner	$\text{CO}_2 - \text{CO}_2$	$\text{CO}_2 - \text{N}_2$	$\text{CO}_2 - \text{He}$	$\left[\frac{1}{\text{sec atm}} \right]$
ν_2	$1.5 \cdot 10^5$	$1.5 \cdot 10^5$	$30.4 \cdot 10^5$	
ν_3	$2.8 \cdot 10^5$	$8.4 \cdot 10^4$	$6.5 \cdot 10^4$	

Tab. 2 Ratenkonstanten des V-T-Transfers der ν_2 - bzw. ν_3 -Normalschwingung bei Stößen mit verschiedenen Partnern /44/.

Für ein Gasgemisch von $\text{CO}_2:\text{N}_2:\text{He} = 3:1:60$ bei 20 atm berechnet man für den V-T-Transfer der ν_3 -Schwingung eine Relaxationszeit $\tau \simeq 700 \text{ nsec}$; für die ν_2 -Schwingung gilt $\tau = 20 \text{ nsec}$. Wie aus Tab. 2 ersichtlich ist, ist die schnelle Entleerung des unteren Laserniveaus eine Folge der Beimischung von He im Lasergas. (Eine weitere Bedeutung von He ist - neben seiner Wirkung als Puffergas zur Erzeugung des hohen Drucks - auch seine Wärmeleitfähigkeit, die etwa 6 mal höher ist als die von N_2 bzw. CO_2 ; He sorgt somit beispielsweise für einen effizienten Transport der durch eine elektrische Gasentladung erzeugten Wärme an die Wandung des Lasers /37/).

Durch den Austausch von Vibrationsenergie bei Stößen zwischen CO_2 -Molekülen (sog. V-V-Transfer) stellt sich jedoch innerhalb einer jeden Normalschwingung ν_i in Zeiten, die viel kleiner sind als die Zeitkonstanten des V-T-Transfers, ein Energiegleichgewicht ein, das sich durch eine sog. Modentemperatur T_i beschreiben läßt. Beispielsweise beträgt die Rate für den Prozeß



$4 \cdot 10^9 \frac{1}{\text{sec atm}} / 45/$. Bei einem Partialdruck von 0.9 atm für CO_2 erhält man als Zeitkonstante 0.3 nsec. Ähnliches gilt für die ν_1 - bzw. ν_2 - Mode. Infolge der in Fermiresonanz befindlichen unteren Laserniveaus erfolgt ein Energieaustausch zwischen der ν_1 - und ν_2 - Normalschwingung mit einer Rate von $7.6 \cdot 10^8 \frac{1}{\text{sec atm}}$, also in einer Zeit von etwa 1.5 nsec für 0.9 atm CO_2 /38/, so daß man für beide Normalschwingungen die gleiche Modentemperatur $T_1 = T_2$ hat.

Die Temperaturen T , $T_1 = T_2$ und T_3 hängen von der Art und Stärke der verwendeten Anregungsmethode ab. Im Falle einer Anregung durch eine elektrische Gasentladung, die durch die ins Gasgemisch injizierte Energiedichte ε (üblicherweise angegeben in $\frac{\text{J}}{\text{atm}}$) und durch das Verhältnis von elektrischer Feldstärke und Druck des Gasgemisches $\frac{E}{p}$ charakterisiert wird (siehe Abschnitt I.1.2), erhält man für Energiedichten $10 \frac{\text{J}}{\text{atm}} < \varepsilon < 250 \frac{\text{J}}{\text{atm}}$ eine kinetische Gastemperatur T zwischen 300 K und 400 K /46/. Die Modentemperaturen $T_1 = T_2$ und T_3 werden von den Größen $\frac{E}{p}$ und ε bestimmt. (Wie im folgenden Abschnitt gezeigt, bestimmt $\frac{E}{p}$ den Bruchteil der Energie, die ins Schwingungssystem der einzelnen Normalmoden injiziert wird.) Die Temperatur T_3 läßt sich aus der Kleinsignalverstärkung bei einer bestimmten Frequenz, z.B. $10P(16)$, berechnen /47/. Für ein Gasgemisch $\text{CO}_2:\text{N}_2:\text{He} = 1:1:18$ betragen die Temperaturen bei $\frac{E}{p} = 6 \frac{\text{kV}}{\text{cm atm}}$ und $\varepsilon \approx 32 \frac{\text{J}}{\text{atm}}$ beispielsweise $T = 315 \text{ K}$, $T_1 = T_2 = 350 \text{ K}$ und $T_3 = 1250 \text{ K}$, d.h. $T_1 \ll T_3$ /47/. Mit Hilfe der Temperaturen T , $T_1 = T_2$ und T_3 läßt sich die Besetzung der Energieniveaus innerhalb jeder Normalschwingung und damit die Kleinsignalverstärkung $\alpha(\nu)$ bestimmen. Dieses Verfahren wird als das Temperaturmodell der Kleinsignalverstärkung bezeichnet. Der Beitrag eines Rotations-Schwingungs-Übergangs zwischen einem oberen Niveau 1 und einem unteren Niveau 2 zur Kleinsignalverstärkung ist nämlich gegeben durch

$$\alpha_{12}(\nu) = \psi_{\text{CO}_2} N \sigma_{12} g(\nu - \nu_{12}) (n_1 - \frac{g_1}{g_2} n_2) \quad (6)$$

Die Terme in Gl.(6) sind dabei wie folgt definiert:

N Anzahl der Moleküle pro Volumeneinheit

ψ_{CO_2} Bruchteil der CO_2 -Moleküle

σ_{12} Absorptionsquerschnitt für den Übergang vom Niveau 1 zu Niveau 2

$g(\nu - \nu_{12})$ Linienformfunktion des Übergangs.

(Sie ist in erster Näherung eine Lorentzfunktion mit einer Breite nach Gl.(5);

Abweichungen treten bei hohen Drücken durch die Wechselwirkung benachbarter Linien auf.)

g_1, g_2 : Entartungsgrad der beiden Niveaus

n_1 und n_2 sind die relativen Besetzungen in den Niveaus 1 und 2 und sind durch die Modentemperaturen T_2 bzw. T_3 gegeben:

$$n_1 \sim \exp(-h\nu_3/kT_3) \cdot \rho(J_1)$$

$$n_2 \sim \exp(-h\nu_{1,2}/kT_2) \cdot \rho(J_2)$$

$\rho(J_i)$ ist dabei die durch die kinetische Gastemperatur T bestimmte relative Besetzung des Rotationsniveaus J_i nach Gl.(3).

Summiert man über alle Rotations-Schwingungs-Übergänge, erhält man die resultierende Kleinsignalverstärkung eines CO₂-Lasers bei der Frequenz ν . Abb. 3 zeigt das Ergebnis einer von Taylor et al. /47/ auf der Basis dieses Temperaturmodells durchgeführten Rechnung für 3 verschiedene Gasdrücke (Mischung CO₂:N₂:He = 1:1:18) bei einer injizierten Energiedichte $\epsilon = 32 \frac{\text{J}}{\text{atm}}$ und $\frac{E}{p} = 6 \frac{\text{kV}}{\text{cm atm}}$. Die dort aufgeführten theoretischen Voraussagen zeigten gute Übereinstimmung mit von Taylor et al. gleichzeitig durchgeführten Messungen der Kleinsignalverstärkung. Man erkennt deutlich, daß für Drücke $p \leq 10$ atm noch eine Modulation der Verstärkung in allen 4 Emissionszweigen des CO₂-Lasers besteht. Bei $p = 19$ atm jedoch zeigen die R -Zweige eine völlig glatte Kleinsignalverstärkung, nur in den P -Zweigen existiert noch eine geringfügige Welligkeit (etwa 7% Modulationstiefe). Dies ist eine Folge des erwähnten größeren Abstandes benachbarter Linien in den P -Zweigen gegenüber dem in den R -Zweigen. Auch verschiebt sich das Maximum der Verstärkung mit zunehmendem Druck vom $10P$ -Zweig in den $9R$ -Zweig, weil dort der Abstand benachbarter Linien am kleinsten ist (vgl. Gl.(2)).

Von Taylor et al. wurden bei der Berechnung der Kleinsignalverstärkung nicht nur das reguläre $9.4 \mu\text{m}$ bzw. $10.4 \mu\text{m}$ Band, sondern auch Beiträge durch die sog. 'sequence-bands' (Übergänge $|00^n\rangle \rightarrow |02^n - 1\rangle, |10^n - 1\rangle, n = 2, 3, 4$) und 'hot-bands' (Übergänge $|01^1 1\rangle \rightarrow |11^1 0\rangle, |01^1 2\rangle \rightarrow |11^1 1\rangle$) berücksichtigt. Denn bei Anregung durch eine elektrische Gasentladung werden die oberen Niveaus dieser Übergänge ebenfalls besetzt (siehe Kap. I.1.2). Ihre Frequenzen liegen zwischen den Linienzentren der regulären Übergänge, so daß ihr Einfluß auf die Verstärkung bei den üblichen CO₂-Lasern, deren Betriebsdruck ≤ 1 atm ist, keine Rolle spielt. Wie in Abb. 3 gezeigt, leisten sie jedoch einen wesentlichen Beitrag zur Kleinsignalverstärkung eines Hochdruck-CO₂-Lasers.

Zum Schluß sei noch kurz auf die Bedeutung von N₂ im Gasgemisch eingegangen. Wie in Abschnitt I.1.2 gezeigt wird, erfolgt eine bevorzugte Anregung von N₂-Schwingungen durch Stöße der freien Elektronen in einer elektrischen Gasentladung mit den N₂-Molekülen.

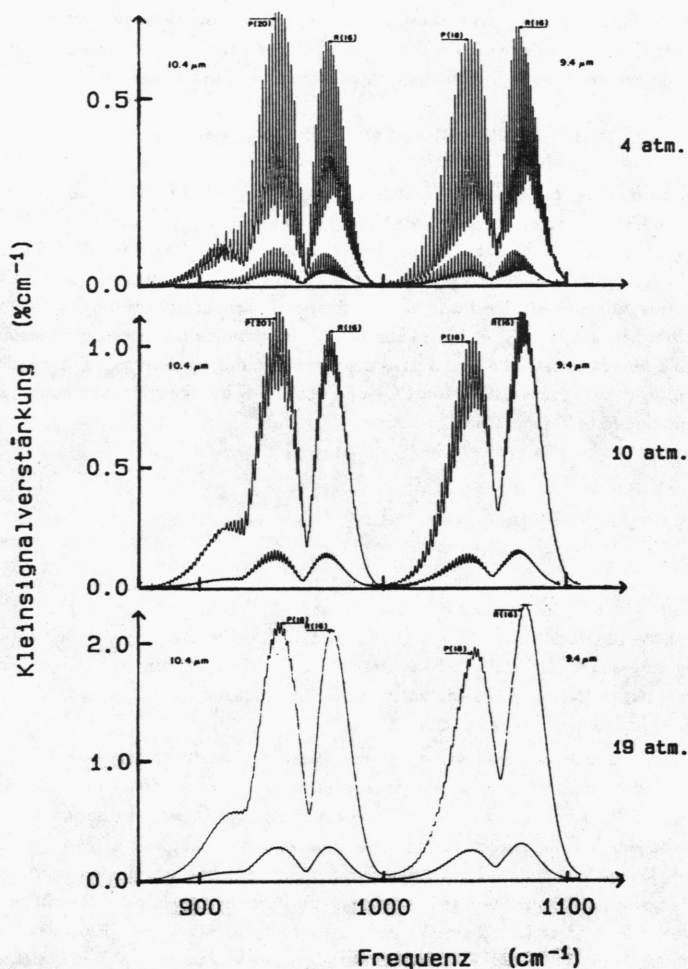


Abb.3 Kleinsignalverstärkung eines Hochdruck- CO_2 -Lasers für einen Betriebsdruck von 4, 10 und 19 atm bei einer injizierten Energiedichte $\varepsilon = 32 \frac{\text{J}}{\text{l atm}}$ und $\frac{E}{p} = 6 \frac{\text{kV}}{\text{cm atm}}$ nach Taylor /47/.
(Die untere Kurve in jedem Spektrum zeigt den Beitrag der 'sequence-bands' zur gesamten Verstärkung.)

Da der erste angeregte Schwingungszustand von N_2 (siehe Abb. 1) energetisch dem $|00^0 1\rangle$ -Zustand des CO_2 -Moleküls sehr nahe liegt, kommt es zu einem Energietransfer zwischen den N_2 -Schwingungen und der ν_3 -Normalschwingung von CO_2 durch Stöße der Form



Der Austausch erfolgt mit einer Rate von $14.7 \cdot 10^6 \frac{1}{\text{sec atm}} / 38$, d.h. bei einem Partialdruck von 0.9 atm CO_2 mit einer Zeitkonstante $\tau_{N_2, CO_2} = 75 \text{ nsec}$. N_2 hat somit die wichtige Bedeutung als Energielieferant und sorgt im wesentlichen für die Besetzungsinversion des CO_2 -Lasers. In Abschnitt I.1.4.2 wird gezeigt, daß die Zeit t_{auf} , in der sich der Laserpuls nach erfolgter Anregung aufbaut, für den in dieser Arbeit entwickelten Hochdruck- CO_2 -Laser etwa 700 nsec beträgt. Da $\tau_{N_2, CO_2} \ll t_{auf}$, kann bei der Berechnung der Kleinsignalverstärkung nach dem Temperaturmodell davon ausgegangen werden, daß ein Energiegleichgewicht zwischen dem ν_3 -Schwingungssystem von CO_2 und dem N_2 -Schwingungssystem vorhanden ist und damit eine feste Temperatur T_3 existiert.

1.2 Anregung eines Hochdruck- CO_2 -Lasers durch elektrische Gasentladungen

Die Anregung durch eine elektrische Gasentladung ist die am weitesten verbreitete Anregungsmethode für Hochdruck- CO_2 -Laser und wird auch bei dem in dieser Arbeit entwickelten Laser eingesetzt. Ziel dieses Abschnitts ist es, die charakteristischen Parameter einer elektrischen Gasentladung klarzustellen und die Bedingungen zu diskutieren, die für eine effiziente Anregung des Lasergases nötig sind.

Auf eine Diskussion der anderen Anregungstechniken von Hochdruck- CO_2 -Lasern, die in den letzten Jahren erprobt wurden, sei hier verzichtet. Eine Übersicht findet man in /35/. Behandelt werden dort insbesondere durch optisches Pumpen angeregte Hochdruck- CO_2 -Laser /19,20/ und durch Hochfrequenzentladungen angeregte Wellenleiterlaser /48/. In beiden Fällen wurde eine kontinuierliche Abstimbarkeit der Laserfrequenz demonstriert. Die erzielten Ausgangspulsleistungen lagen jedoch nur im Bereich von einigen Kilowatt und sind damit, wie in Abschnitt I.2.1.2 gezeigt wird, zu gering, um stimulierte Ramanemission in CH_3F hervorzurufen. Hierfür werden Pulsleistungen im Megawattbereich benötigt. Diese können nur von durch elektrische Gasentladungen angeregten Hochdruck- CO_2 -Lasern geliefert werden.

1.2.1 Prinzip der elektrischen Gasentladung

Abb. 4 zeigt die prinzipielle Anordnung, die üblicherweise für die Anregung von Lasern durch gepulste elektrische Gasentladungen verwendet wird. Ein Kondensator C wird über einen Ladewiderstand R_L bei offenem Schalter S auf eine Spannung U_C aufgeladen, die über der Durchbruchspannung des Gasgemisches zwischen den Elektroden der Entladungsstrecke liegt. Nach Schließen des Schalters S (realisiert durch eine getriggerte Funkenstrecke) liegt schlagartig die volle Spannung U_C des Kondensators zwischen den Elektroden an. Eine anfänglich geringe Zahl freier Elektronen wächst dann durch Stoßionisationslawinenprozesse (Avalancheprozesse) exponentiell an. Der Strom in der Entladung nimmt zu, der Spannungsabfall an den Elektroden infolge des bekanntlich negativen differentiellen Widerstands der Spannungs-Strom-Charakteristik $U = U(I)$ einer Gasentladung ab. Die Stromzunahme kann jedoch nicht beliebig weit erfolgen, da die endliche Impedanz R_E des Entladekreises - sie wird gebildet aus dem ohmschen Widerstand, der Induktivität und Kapazität von Kabeln und Funkenstrecke - den Strom begrenzt. Es stellt sich ein stabiler Arbeitspunkt ein, der sich aus dem Schnittpunkt der Spannungs-Strom-Charakteristik $U = U(I)$ der Entladungsstrecke mit der Arbeitsgeraden, die von der Kondensatorspannung U_C und der Impedanz R_E bestimmt wird, ergibt (Abb. 4). Die Entladung verhält sich ab diesem Zeitpunkt wie eine Gleichstromgasentladung, wobei jedoch die angelegte Versorgungsspannung U_C infolge der Entladung des Kondensators mit der Zeit abnimmt.

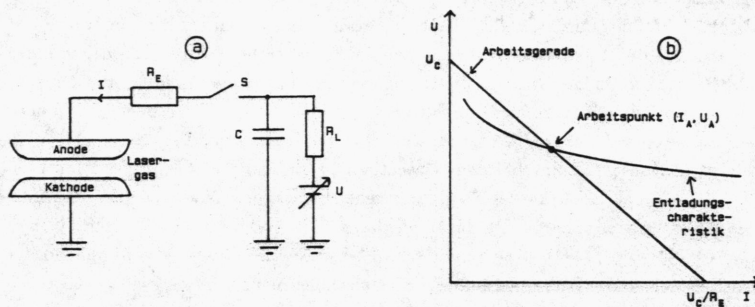


Abb. 4 a) Prinzipielle Anordnung für eine gepulste Gasentladung.
b) Gleichgewichtsbedingung für den Entladestrom I_A .

Werden während des Entladungsprozesses die Verluste an freien Elektronen innerhalb der Entladungsstrecke (z.B. durch Rekombination der Elektronen mit positiv geladenen Ionen, durch Elektroneneinfang durch neutrale Moleküle oder durch Diffusion der Elektronen aus dem Entladungsvolumen) durch die Lawinenprozesse ausgeglichen, spricht man von einer selbständigen Gasentladung (Townsend-Avalanche-Entladung). Muß dagegen der Elektronenverlust durch einen extern kontrollierten Erzeugungsprozeß von Elektronen (z.B. mit Hilfe von Elektronen-, UV- oder Röntgenstrahlung) kompensiert werden, heißt die Entladung unselbständig. Beide Entladungstypen werden für die Anregung von Hochdruck-CO₂-Lasern verwendet und sollen im folgenden kurz diskutiert werden. Dabei wird jedoch nicht auf die grundlegende Theorie von Gasentladungen eingegangen - Näheres dazu findet man in dem Buch von Nasser /49/, sondern vielmehr wird dargelegt, welche elektrischen Feldstärken bei einem gegebenen Gasmisch CO₂:N₂:He und gegebenem Druck p für die Aufrechterhaltung einer selbständigen Gasentladung nötig sind und wie sich dann die Energie der im elektrischen Feld der Gasentladung beschleunigten Elektronen durch inelastische Stöße auf einzelne Anregungen der im Gasmisch enthaltenen Moleküle (z.B. auf Schwingungsanregungen, auf elektronische Anregungen etc.) verteilt. Die optimalen Anregungsbedingungen lassen sich daraus ermitteln.

1.2.2 Bedingungen für die effiziente Anregung des Lasergases in einer elektrischen Gasentladung

Ausgangspunkt aller Überlegungen für die optimalen Anregungsbedingungen ist die Geschwindigkeitsverteilungsfunktion $f(v)$ der Elektronen in der Gasentladung. Sie wurde von Lowke et al. /50/ durch numerische Lösung der Boltzmannschen Transportgleichung unter Berücksichtigung von 31 inelastischen Streuprozessen der Elektronen - die wichtigsten sind aus Abb. 6 ersichtlich - für verschiedene Gasmischungen CO₂:N₂:He berechnet. Der fundamentale Parameter, von dem die Verteilungsfunktion abhängt, ist hierbei das Verhältnis $\frac{E}{N}$ von elektrischer Feldstärke innerhalb der Gasentladung und der Teilchendichte N im Gas. (Statt $\frac{E}{N}$ wird oftmals auch die Größe $\frac{E}{p}$ angegeben. Zwischen N und p (p = Gesamtdruck des Gases) gilt annähernd der Zusammenhang $N = \frac{p}{kT}$; für $T = 300$ K erhält man $N = 2.4 \cdot 10^{19} \frac{\text{cm}^{-3}}{\text{atm}} \cdot p / 44$). Die physikalische Bedeutung der Größe $\frac{E}{N}$ wird anhand des folgenden Beispiels klar: Tritt innerhalb der Gasmischung nur ein inelastischer Stoßprozeß der Elektronen mit dem Wirkungsquerschnitt σ auf, erhält man mit Hilfe der mittleren freien Weglänge $\lambda = \frac{1}{N\sigma}$ der Elektronen die Energie W , die ein Elektron zwischen 2 Stößen aufnimmt, aus der Bedingung $W = eE\lambda = \frac{e}{\sigma}(\frac{E}{N})$ (e = Elementarladung). Die Größe $\frac{E}{N}$ ist somit ein Maß für die mittlere Energie der stoßenden Elektronen. Kennt man die Geschwindigkeitsverteilung der Elektronen (die i.a. keine Maxwellsche Geschwindigkeits-

verteilung ist /50/) in einer gegebenen Gasmischung bei einem gegebenen Wert $\frac{E}{N}$, lassen sich daraus die wesentlichen Transportkoeffizienten berechnen, die die makroskopischen Eigenschaften der Entladung unter den gegebenen Bedingungen bestimmen. Dies sind vor allem die Driftgeschwindigkeit v_D der Elektronen, der sog. erste Townsendsche Ionisationskoeffizient α , der über $dn_e = \alpha n_e dx$ die Zunahme dn_e einer Elektronenkonzentration n_e durch Stoßionisationsprozesse längs der Wegstrecke dx beschreibt, und der sog. Einfangkoeffizient a , der die Abnahme einer Elektronenkonzentration längs einer Wegstrecke infolge des Elektroneneinfangs durch neutrale Teilchen im Gas beschreibt. (Der wichtigste Einfangprozeß für Elektronen ist der sog. dissoziative Einfang $\text{CO}_2 + e^- \rightarrow \text{CO} + \text{O}^-$ /51/; jedoch treten auch die direkten Einfänge der Form $\text{M} + e^- \rightarrow \text{M}^-$ auf, wobei M Moleküle wie CO_2 , N_2 aber auch Folgeprodukte von Dissoziationsprozessen [siehe Anmerkung am Ende dieses Abschnitts] bezeichnet). Diese Transportkoeffizienten sind durch ihre Abhängigkeit von der Geschwindigkeitsverteilungsfunktion auch eine Funktion von $\frac{E}{N}$. Nach /50/ gilt für die Elektronenkonzentration in der Entladungsstrecke die Kontinuitätsgleichung

$$\frac{dn_e}{dt} = n_e \alpha v_D - n_e a v_D - \gamma n_e^2 + S_{ext} \quad (8)$$

αv_D ist, wie bereits erwähnt, die Ionisierungsrate von neutralen Gasmolekülen durch die Elektronen, $a v_D$ die Einfangrate freier Elektronen durch neutrale Teilchen; γn_e^2 beschreibt die Rekombination von Elektronen und positiven Ionen - die Dichte von Elektronen und positiven Ionen sei als gleich angenommen! -, S_{ext} die Erzeugung von Elektronen im Gas durch externe Einflüsse (z.B. Elektronenstrahl, UV-Strahlung etc.). Im Falle eines konstanten Stromflusses in der Entladung (Arbeitspunkt!) muß aus Gründen der Ladungserhaltung gelten $\frac{dn_e}{dt} = 0$ /50/. Unter Verwendung der Stromdichte $j = n_e e v_D$ erhält man dann für den Fall $S_{ext} = 0$:

$$\frac{j}{N} = \frac{ev_D^2}{\gamma} \left(\frac{\alpha}{N} - \frac{a}{N} \right) \quad (9)$$

Damit eine selbständige Gasentladung aufrechterhalten wird, muß gelten:

$$j \geq 0, \quad \text{d.h.} \quad \frac{\alpha}{N} \geq \frac{a}{N} \quad (10)$$

In Abb. 5 sind die von Lowke et al. /50/ berechneten Größen $\frac{\alpha}{N}$ und $\frac{a}{N}$ in Abhängigkeit von $\frac{E}{N}$ für 3 verschiedene Gasmischungen aufgetragen. Bei der Bestimmung von $\frac{\alpha}{N}$ wurden die direkten Ionisationsprozesse von CO_2 und N_2 , aber auch die sog. Penningionisation /50/ durch angeregte Heliumzustände berücksichtigt. Bei $\frac{a}{N}$ wurde nur der hauptsächliche Elektroneneinfangprozeß, nämlich der dissoziative Einfang $\text{CO}_2 + e^- \rightarrow \text{CO} + \text{O}^-$ berücksichtigt.

Das für die Aufrechterhaltung einer selbständigen Entladung nötige Mindestverhältnis $\frac{E}{N}$ läßt sich damit aus Abb. 5 und Gl.(10) bestimmen. Man erhält für die dort angegebenen

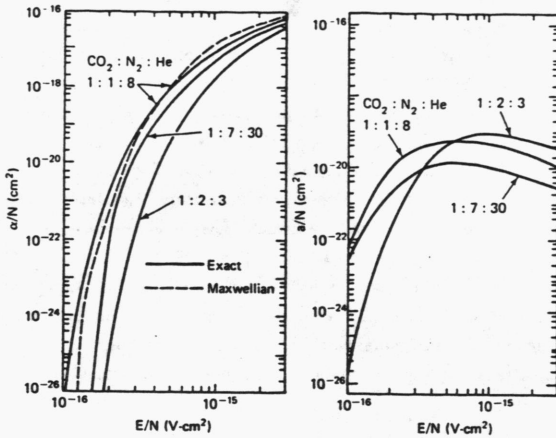


Abb. 5 Abhängigkeit des Ionisierungskoeffizienten α und des Einfangkoeffizienten a von $\frac{E}{N}$ für verschiedene Gasmischungen $\text{CO}_2:\text{N}_2:\text{He}$ nach Lowke /50/.

Gasmischungen folgende Werte:

Mischung $\text{CO}_2:\text{N}_2:\text{He}$	$\frac{E}{N}$ [10^{-16}V cm^2]
1:2: 3	4.9
1:1: 8	2.8
1:7:30	2.7

Tab. 3 Mindestverhältnis $\frac{E}{N}$ für Aufrechterhaltung einer selbständigen Gasentladung für verschiedene Gasmischungen.

Mit zunehmender He-Konzentration nimmt somit die nötige Feldstärke E bei einem gegebenen festen Gesamtdruck p ab. Ursache ist, daß die Elektronen kaum inelastische Stoßprozesse an Heliumatomen erleiden (die niederenergetischste Anregung eines He-Atoms liegt bei etwa 20 eV; für ein Gasgemisch 1:7:30 beträgt der Wert der Energieverteilungsfunktion der Elektronen bei 10 eV nur mehr 1% des Maximalwertes bei 2 eV /50/). Die Elektronen besitzen also bei einem gegebenen Gasdruck p im Falle eines hohen He-Anteils eine höhere mittlere Energie als bei geringem He-Anteil. Die Verteilungsfunktion verschiebt sich zu höheren Energien hin. Für die Gasmischung $\text{CO}_2:\text{N}_2:\text{He} = 3:1:60$, die in dem in dieser Arbeit beschriebenen Laser verwendet wurde, kann somit ein nötiges $\frac{E}{N}$ von etwa $2.5 \cdot 10^{-16} \text{ V cm}^2$

abgeschätzt werden. Bei einem Druck von 20 atm und einem Elektrodenabstand von 1 cm muß daher die Spannung an den Elektroden im Arbeitspunkt etwa 120 kV betragen.

Aus der Geschwindigkeitsverteilung der Elektronen in der Gasentladung läßt sich auch berechnen, wie sich die Energie der Elektronen durch inelastische Stöße auf einzelne Anregungen der Gasmoleküle verteilt. Abb.6 zeigt das Ergebnis einer solchen Rechnung, die von Nighan et al. /52/ für eine Gasmischung $\text{CO}_2:\text{N}_2:\text{He} = 1:1:8$ durchgeführt wurde. Eingezeichnet sind nur der auf elektronische Anregungen von CO_2 und N_2 , der auf die Anregung der $|00^\circ 1\rangle$ bzw. $|01^1 0\rangle$ Schwingungszustände von CO_2 sowie der auf die Schwingungszustände $|v\rangle v = 1 \dots 8$ von N_2 übertragene Bruchteil der Elektronenenergie. Bei den Schwingungsanregungen von N_2 ist jeweils der Beitrag der Zustände mit $v = 1 \dots 4$ bzw. $v = 5 \dots 8$ zu einer Kurve zusammengefaßt.

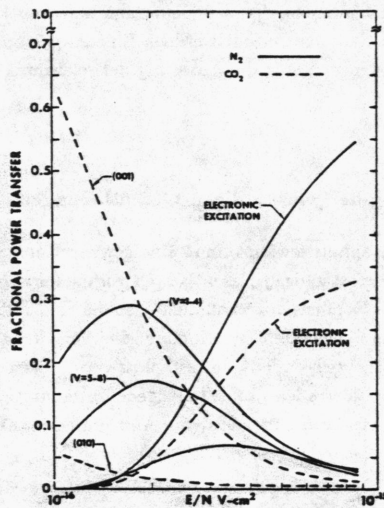


Abb.6 Bruchteile der an verschiedene Anregungen von CO_2 - und N_2 -Molekülen übertragenen Elektronenenergie als Funktion von $\frac{E}{N}$ für eine Gasmischung $\text{CO}_2:\text{N}_2:\text{He} = 1:1:8$ nach Nighan /52/.

Für $\frac{E}{N} = 2.5 \cdot 10^{-16} \text{ V-cm}^2$ werden somit etwa 35% der Elektronenenergie auf die Anregung von N_2 -Schwingungen, 15% auf die Anregung der $|00^\circ 1\rangle$ -Schwingung von CO_2 ver-

wandt. Für $\frac{E}{N} = 4 \cdot 10^{-16} \text{ V cm}^2$ sind dies nur mehr 20% bzw. 6%; dagegen werden in diesem Fall 55% der Elektronenenergie in elektronische Anregungen umgesetzt. Rechnungen für Gasgemische mit höheren He-Anteilen wurden von Nighan nicht durchgeführt. Da jedoch eine Erhöhung des He-Anteils bei festem Verhältnis $\frac{E}{N}$ eine Erhöhung der mittleren Elektronenenergie bewirkt, muß in diesem Fall eine Verschiebung der Kurven in Abb. 6 zu geringeren $\frac{E}{N}$ -Werten hin erfolgen. Für eine Gasmischung $\text{CO}_2:\text{N}_2:\text{He} = 3:1:60$ und einen Wert $\frac{E}{N} = 2.5 \cdot 10^{-16} \text{ V cm}^2$ läßt sich somit eine Umsetzung der Elektronenenergie in die Anregung der relevanten Schwingungszustände von CO_2 und N_2 von weniger als 20% vermuten. Berücksichtigt man zusätzlich den Quantenwirkungsgrad des CO_2 -Lasers, d.h. das Verhältnis der Anregungsenergie der $|00^0 1\rangle$ -Schwingung zur Energie der Laserübergänge, das etwa 40% beträgt, erhält man für das gegebene Beispiel einen theoretischen Energiekonversionskoeffizienten von 8%.

Hier wird dann der Vorteil von unselbständigen Gasentladungen sichtbar. In diesem Fall ist in Gl.(8) $S_{ext} \neq 0$. Die Bedingung $\frac{dn_a}{dt} = 0$ kann dann bereits bei geringeren $\frac{E}{N}$ -Werten als bei selbständigen Gasentladungen erfüllt werden. Eine Anpassung von $\frac{E}{N}$ für die optimale Umsetzung der Elektronenenergie in die Anregung der relevanten Molekülschwingungen ist dann möglich.

1.2.3 Vorionisierung bei selbständigen Gasentladungen

Wie zu Beginn dieses Kapitels erwähnt, muß eine gewisse Konzentration freier Elektronen vor dem Zünden einer selbständigen Gasentladung vorhanden sein, damit der Entladungsprozeß nach Anlegen der Kondensatorspannung U_C an die Elektroden auch einsetzt. Ist diese Konzentration zu gering oder besteht eine inhomogene Verteilung der freien Elektronen im Entladungsvolumen, bildet sich - vor allem bei hohen Gasdrücken und dementsprechend hohen Spannungen an den Elektroden - keine homogene Entladung im gesamten Volumen zwischen den Elektroden aus, was für eine effiziente Anregung des Lasergases erforderlich wäre, sondern es kommt zur Ausbildung einer sog. Bogenentladung [53]. Die mikroskopischen Prozesse, die zum Auftreten einer solchen Bogenentladung führen, sind relativ kompliziert und sollen hier nicht näher diskutiert werden. Eine ausgezeichnete Behandlung dieser Problematik findet man in [53,54]. Es sei nur erwähnt, daß sich bei einer Bogenentladung ein eng begrenzter Kanal (mit einem typischen Durchmesser von einigen mm) aus Plasma, d.h. einem Gemisch von Elektronen und positiven Ionen, bildet, der die Elektroden kurzschließt. Die gesamte Ladung des Kondensators fließt über diesen Plasmakanal ab. Eine Anregung der Gasmoleküle erfolgt deswegen nur innerhalb des Kanals. Auch sind Bogenentladungen wegen ihrer zerstörerischen Wirkung auf die Elektroden wenig wünschenswert.

Bogenentladungen können vermieden werden, wenn vor dem Zünden der Entladung für eine

genügend große und homogen im Entladungsvolumen verteilte Konzentration von freien Elektronen gesorgt wird. Nach Karyushin /55/ muß für die Mindestkonzentration n_{eo} dieser Elektronen bei einem $\frac{E}{N}$ -Wert von $3 \cdot 10^{-16} \text{ V cm}^2$ gelten:

$$n_{eo} \geq 10^4 \text{ cm}^{-3} \text{ atm}^{-3} \cdot p^3 \quad (11)$$

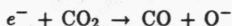
Für $p = 20 \text{ atm}$ ergibt dies $n_{eo} \geq 10^8 \text{ cm}^{-3}$.

Eine erprobte Methode, Elektronenkonzentrationen dieser Größenordnung zu erzeugen, ist die Vorionisierung des Gases im Raum zwischen den Elektroden mit Hilfe von UV-Strahlung. Diese kann beispielsweise durch das Zünden einer Reihe von zusätzlichen Funkenstrecken, die parallel zu den Elektroden an beiden Seiten des Entladungsvolumens angebracht sind, erzeugt werden. McKen et al. /56/ haben gezeigt, daß hierbei infolge der Rekombination von freien Elektronen mit ionisierten N_2 -Molekülen in den Funkenstrecken harte UV-Strahlung im Bereich von 100 nm - 240 nm emittiert wird. Die Effizienz der Vorionisierung durch UV-Strahlung wird jedoch dadurch begrenzt, daß Strahlung der Wellenlänge $\leq 170 \text{ nm}$ durch CO_2 -Moleküle stark absorbiert wird. (Die Eindringtiefe bei einem Partialdruck von 0.9 atm für CO_2 beträgt weniger als 1 mm /57/). Nur die niederenergetischeren UV-Photonen können demnach merklich ins Entladungsvolumen eindringen. Ihre Energie reicht jedoch für die direkte Ionisierung der Moleküle im Gasgemisch nicht mehr aus /38/, die Erzeugung freier Elektronen erfolgt nunmehr über die Ionisierung von CO_2 , N_2 oder He durch Zwei- oder Mehrphotonenprozesse. Erhöhen läßt sich die Ionisierungsrate jedoch, wenn man Moleküle mit geringer Ionisationsenergie dem Lasergasgemisch beimischt. Besonders geeignet sind Trimethylamin oder Tripropylamin, die durch Strahlung der Wellenlänge $\geq 170 \text{ nm}$ bereits ionisiert werden können.

Die Methode der UV-Vorionisierung mittels Funkenstrecken wurde auch bei dem in dieser Arbeit entwickelten Hochdruck- CO_2 -Laser verwendet. (Die erwähnte besondere Anordnung und ihre Vorteile gegenüber Anordnungen, wie sie z.B. von Taylor et al. /47/ beschrieben wurden, werden im folgenden Abschnitt diskutiert.) Einen Überblick über die verschiedensten technischen Ausführungen von UV-vorionisierten Hochdruck- CO_2 -Lasern mit selbständiger Gasentladung findet man in /35,37,38/. Vor kurzem wurde von Wan et al. auch eine selbständige Gasentladung in Hochdruck- CO_2 -Lasern mit einer Elektronenstrahlvorionisierung demonstriert /58/. Mit Elektronenstrahlvorionisierung erhält man leichter die gewünschte Konzentration freier Elektronen im Entladungsvolumen als mit UV-Vorionisierung. Absorptionsverluste, wie sie bei letzterer diskutiert wurden, treten nicht auf. Dafür steigt der technische Aufwand beträchtlich /58/. Konkrete Daten, wie beispielsweise Anordnung der UV-Vorionisierung, $\frac{E}{N}$ -Werte, injizierte Energiedichte ϵ und erzielte Ausgangspulsenergien von 3 repräsentativen Beispielen von UV-vorionisierten Hochdruck- CO_2 -Lasern mit selbständiger Gasentladung sind in Abschnitt I.1.5 zu finden. Dort wird auch ein Vergleich mit den Daten des in dieser Arbeit entwickelten Lasers durchgeführt.

Auf die Diskussion der Eigenschaften, der Vor- und Nachteile von Lasern, die durch nicht-selbständige Gasentladungen angeregt werden, sei hier verzichtet. Informationen finden sich in /21-24/.

Anmerkung: Ein Problem, das bei durch elektrische Gasentladungen angeregten CO₂-Lasern auftritt, sei noch angeschnitten, nämlich das Problem der Dissoziation von CO₂-Molekülen in der elektrischen Entladung aufgrund der Prozesse



Sie machen den langfristigen Betrieb von 'sealed off' -Hochdruck-CO₂-Lasern unmöglich. Zum einen wird durch die Dissoziation die Konzentration der CO₂-Moleküle in der Gas Mischung verringert. Die Ausgangspulsenergie wird daher bei fester injizierter Energiedichte ϵ mit zunehmender Zahl der Entladungen abnehmen. Zum anderen entstehen aus den Komponenten CO und O durch Zwei- oder Dreiteilchenstöße untereinander, mit N₂-Molekülen oder Elektronen diverse Folgeprodukte wie O₂, N₂O oder NO. Der dissoziative Einfang freier Elektronen durch diese Folgeprodukte ist um einen Faktor 10 - 100 wahrscheinlicher als der dissoziative Einfang durch CO₂ /59/. Eine zunehmende Konzentration der Folgeprodukte bewirkt damit einen verstärkten Einfang der durch die UV-Vorionisierung erzeugten freien Elektronen. Die Wahrscheinlichkeit für Bogenentladungen nimmt zu. Eine Rückgewinnung von CO₂ aus CO und O₂ ist jedoch möglich, wenn man das Gasgemisch durch einen kommerziellen Platin- oder Hopkalit-Katalysator /60/, der sich innerhalb des geschlossenen Systems befindet, pumpt.

1.3 Aufbau des Hochdruck-CO₂-Lasers

Der im Rahmen dieser Arbeit entwickelte Hochdruck-CO₂-Laser /61,62/ enthält zwei wesentliche Neuerungen gegenüber den bisher veröffentlichten Konstruktionen von UV-vorionisierten Hochdruck-CO₂-Lasern mit selbständiger Gasentladung /17,25-29,63/.

- 1) Eine neuartige Anordnung für die UV-Vorionisierung ermöglichte eine besonders einfache und kompakte Laserkonstruktion und einen dauerhaften Betrieb ohne Bogenentladungen auch bei Gasdrücken $p \geq 20$ atm. (Dies sind - nebenbei bemerkt - die höchsten Gasdrücke, die bei UV-vorionisierten Hochdruck-CO₂-Lasern verwendet wurden. So betrug der typische Betriebsdruck von Lasern, bei denen eine kontinuierliche Abstimmöglichkeit demonstriert wurde, 10-12 atm /17,27,63/. Der höchste Betriebsdruck $p = 19$ atm wird von Taylor et al. /47/ berichtet; Emissionscharakteristiken bei diesem Druck wurden von ihnen jedoch nicht veröffentlicht.)

- 2) Die Verwendung von 2 Entladungsstrecken ergab eine deutliche Ausweitung des Abstimmbereichs und eine deutliche Erhöhung der Ausgangspulsenergie bei gleichzeitiger Reduktion der Versorgungsspannung gegenüber einem Betrieb mit nur einer Entladungsstrecke, wie er bisher üblich war.

Ziel dieses Abschnitts ist es nun, die wesentlichen technischen Bauteile dieses Lasers zu beschreiben.

1.3.1 Elektroden

Eine wichtige Voraussetzung für homogene elektrische Entladungen ist neben einer genügend starken Vorionisierung natürlich eine Elektrodenform, die zu einer möglichst homogenen elektrischen Feldverteilung im Entladungsvolumen führt (sog. 'uniform field'-Elektroden). Sehr häufig wird ein Rogowski-Profil /37/ für die Elektrodenform gewählt. Dieses besitzt den Nachteil, daß die Breite der Elektroden viel größer als ihr Abstand sein muß, damit die elektrische Feldverteilung zwischen den Elektroden eine gute Homogenität besitzt /37/. Der Platzbedarf für die Elektroden wird sehr groß. Für den hier beschriebenen Laser wurde daher ein anderes Profil gewählt, das von Chang /64/ berechnet wurde und sehr kompakte Elektrodenformen gestattet.

Im folgenden mögen x, y die kartesischen Koordinaten in der Querschnittsebene der Elektroden bezeichnen, wobei der Nullpunkt in der Mitte zwischen beiden Elektroden liegen möge (vgl. Abb. 8). Wie aus der Funktionentheorie bekannt ist, erfüllen bei jeder umkehrbaren konformen Abbildung

$$f: C \rightarrow C : f(x + iy) = u(x, y) + iv(x, y)$$

die reellen Funktionen $u, v: \mathbb{R}^2 \rightarrow \mathbb{R}$ die Laplacegleichung der Elektrostatik. $v(x, y)$ läßt sich daher beispielsweise als das elektrostatische Potential am Ort (x, y) und $u(x, y)$ als die zugehörige Flußfunktion an diesem Ort interpretieren /64/. Damit wird durch die Vorgabe einer beliebigen konformen Abbildung f ein Potential und dadurch eine elektrische Feldverteilung $\vec{E}(x, y)$ in der Ebene definiert. Statt f kann man sich auch die Umkehrabbildung $x + iy = f^{-1}(u + iv)$ vorgeben.

Chang /64/ hat folgende konformale Abbildung vorgeschlagen:

$$x + iy = (u + iv) + k \sinh(u + iv) \quad k \in \mathbb{R}, k > 0 \quad (12)$$

Da die Elektrodenoberfläche eine Äquipotentialfläche ist, muß hier für sie gelten: $v(x, y) = v_0 = \text{konst.}$ Damit wird das Elektrodenprofil durch die folgende Kurve beschrieben:

$$\begin{aligned} x &= u + k \cos v_0 \sinh u \\ y &= v_0 + k \sin v_0 \cosh u \end{aligned} \quad u \in \mathbb{R} \text{ bel.} \quad (13)$$

Nach Chang muß für eine möglichst flache Verteilung des Betrags E des elektrischen Feldes in der Mitte der Elektroden (d.h. bei $u = 0$) gelten: $\frac{\partial E^{-2}}{\partial u^2} \Big|_{u=0} = 0$. Da $E^{-2} = \left| \frac{\partial(x+iy)}{\partial(u+iv)} \right|^2$, erhält man die Bedingung

$$v_0 = \arccos(-k) \quad (14)$$

In Abb. 7 sind die normierten Profile im 1. Quadranten für drei verschiedene k -Werte aufgetragen. Wie man dort sieht, bestimmt der k -Wert die Ausdehnung der Elektroden. Um den k -Wert festzulegen, gibt man sich ein Intervall $[-x_m, x_m]$ (dieses entspricht eindeutig einem Intervall $[-u_m, u_m]$) um das Zentrum der Elektroden bei $x = 0$ (bzw. $u = 0$) vor, in dem die relative Änderung des Betrags des elektrischen Feldes einen gewünschten Wert δ_m nicht überschreiten darf, d.h.

$$\left| \frac{E(0) - E(u)}{E(0)} \right| \leq \delta_m \quad \forall u \in [-u_m, u_m] \quad (15)$$

Chang gibt folgende Näherungsformel an, um den k -Wert aus den gewünschten Größen x_m, δ_m und y_0 (= Höhe der Elektrode bei $x = 0$; diese entspricht dem halben gewünschten Abstand der beiden Elektroden) zu berechnen.

$$\left(\frac{\sqrt[3]{k}(1 - 0.64\sqrt[3]{k})}{1 - 0.64\sqrt[3]{k} - 0.36\sqrt[3]{k^2}} \right)^6 \left(\cosh \frac{\pi x_m}{2 y_0} - 1 \right)^2 = \frac{1}{(1 - \delta_m)^2} - 1 \quad (16)$$

Für $y_0 = 0.5$ cm, $x_m = 0.55$ cm und $\delta_m = 1\%$ erhält man $k = 0.06$. Dies entspricht der Kurve B in Abb. 7.

Ein etwas kompakteres Elektrodenprofil erhält man, wenn man statt Gl.(14) $\cos v_0 < -k$ wählt. Dann ist die Verteilung $E(u)$ nicht mehr maximal flach bei $u = 0$ und besitzt lokale Maxima bei $u_d = \pm \operatorname{arccosh}\left(-\frac{\cos v_0}{k}\right)$; bei $u = 0$ ist ein lokales Minimum. Definiert man $\delta_d = \left| \frac{E(u_d) - E(0)}{E(0)} \right|$, gilt der Zusammenhang

$$(1 + \delta_d)^2 - 1 = \frac{k + \cos v_0}{(1 - k^2)(1 - \cos^2 v_0)} \quad (17)$$

Solange $\delta_d < \delta_m$ ist, gilt nach wie vor für $|u| \leq |u_d|$ die Bedingung (15), für $\delta_d \rightarrow 0$ ergibt sich wieder (14). Um ein solches Profil zu entwerfen, muß man sich also noch δ_d vorgeben. Damit läßt sich aus (17) v_0 bestimmen. Mittels (13) erhält man das Elektrodenprofil. Im vorliegenden Fall wurde $\delta_d = 0.5\%$ gewählt. Man erhält $v_0 = 1.73$, $u_d = 1.62$ und $x_d = 0.43$ cm. Das resultierende Profil ist die gestrichelte Kurve B' in Abb. 7.

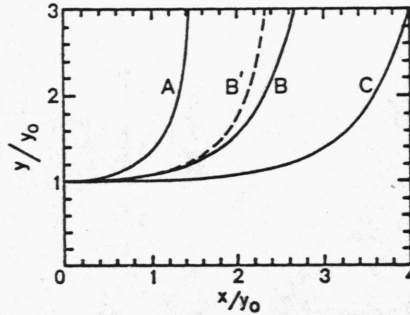


Abb. 7 Normiertes Elektrodenprofil nach Chang /64/ für 3 verschiedene k -Werte: $k = 0.2$ (A), $k = 0.06$ (B), $k = 0.01$ (C).

Die Kurve B' ist das mit $\delta_d = 0.5\%$ kompaktifizierte Profil von B.

Die zur Berechnung des Elektrodenprofils verwendeten Parameter seien noch einmal kurz zusammengestellt:

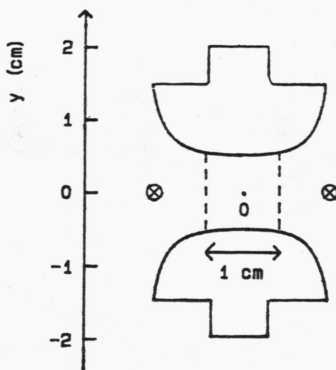
$$\left. \begin{array}{ll} \delta_m = 1\% & 2y_0 = 1 \text{ cm} \\ \delta_d = 0.5\% & 2x_m = 1.1 \text{ cm} \end{array} \right\} k = 0.06 \quad v_0 = 1.73 \quad 2x_d = 0.86 \text{ cm}$$

Als Elektrodenmaterial wurde Aluminium gewählt. Die mechanische Bearbeitung erfolgte mit einem speziell angefertigten, dem Elektrodenprofil entsprechenden Fräser. Die Länge der Elektroden betrug 40 cm, die Dicke 1 cm ($= 2 y_0$), die entsprechende Ausdehnung in x -Richtung 1.18 cm ($= 2.36 y_0$; siehe Abb. 7). An den Enden wurden die Elektroden mittels des Fräasers mit einem Radius von 1.18 cm entsprechend der Ausdehnung der Elektroden rundgefräst. (Man beachte, daß die oben skizzierte Berechnung des Elektrodenprofils ja nur das zweidimensionale Profil in der Querschnittsfläche der Elektroden war.) Sehr wichtig war es, beim Einbau eines Elektrodenpaares darauf zu achten, daß der Abstand der Elektroden auf die ganze Länge auf mindestens 0.1 mm genau war, da sonst die Wahrscheinlichkeit für Bogenentladungen zunahm. Aus dem gleichen Grund erwies es sich als vorteilhaft, die Elektroden sandzustrahlen statt zu polieren.

In Abb. 8 ist noch einmal maßstabsgerecht der Querschnitt durch ein Elektrodenpaar gezeigt. Eingezeichnet ist auch der Querschnitt des beobachteten Entladungsvolumens. Er beträgt etwa $1 \times 1 \text{ cm}^2$. Dies ist der Bereich, innerhalb dem $\left| \frac{E-E(0)}{E(0)} \right| \leq 1\%$ ist. Das gesamte Entladungsvolumen einer Entladungsstrecke, von denen zwei im Laser enthalten waren, hat somit die Größe $1 \times 1 \times 40 \text{ cm}^3$.

Abb. 8

Querschnitt durch ein Elektrodenpaar mit angedeutetem Entladungsvolumen. Die Markierungen \otimes kennzeichnen die Position der Vorionisierungsfunknestrecken.



1.3.2 Vorionisierung

Die Vorionisierung des Lasergasgemisches erfolgte mit Hilfe von Funkenstrecken, die zu beiden Seiten des Entladungsvolumens in der Mitte zwischen beiden Elektroden angebracht waren. Jede Funkenstrecke bestand aus einem Paar von zugespitzten Wolframnadeln (Länge 5 mm, \varnothing 0.5 mm) mit einem Abstand von 2 mm. Der seitliche Abstand von der Mitte des Entladungsvolumens war 1.2 cm (vgl. Abb. 8). Die Wolframnadeln wurden von Messingplättchen gehalten, die ihrerseits auf seitlich des Entladungsvolumens angebrachten Plexiglasblöcken montiert waren. Eine Prinzipskizze des Aufbaus zeigt Abb. 9. Die Anzahl der Funkenstrecken auf den beiden Seiten des Entladungsvolumens betrug 16 bzw. 17 Stück. Die beiden Reihen waren leicht zueinander versetzt, um eine möglichst homogene Ausleuchtung des Entladungsvolumens mit UV-Strahlung zu erreichen.

Das elektrische Anschlußschema zeigt Abb. 10. Alle Funkenstrecken waren parallel geschaltet und über Kondensatoren mit den Elektroden der Hauptentladungsstrecke gekoppelt. Bei den verwendeten Kondensatoren handelte es sich um Keramik Kondensatoren der Kapazität 140 pF und einer Spannungsfestigkeit von 40 kV. Da innerhalb des Kreises einer Funkenstrecke 4 seriell geschaltete Kondensatoren enthalten waren, konnten Spannungen bis maximal 160 kV an den Hauptelektroden anliegen, ohne daß die Durchschlagspannung der Kondensatoren überschritten wurde. Aufgabe der Kondensatoren war es, zu verhindern, daß eine Vorionisierungsfunknestrecke den gesamten Entladekreis kurzschließt. Die benachbarten Nadeln von 2 nebeneinanderliegenden Funkenstrecken waren, wie aus Abb. 9 und 10 ersichtlich, auf gleichem Potential, um einen Überslag zwischen den nebeneinanderliegen-

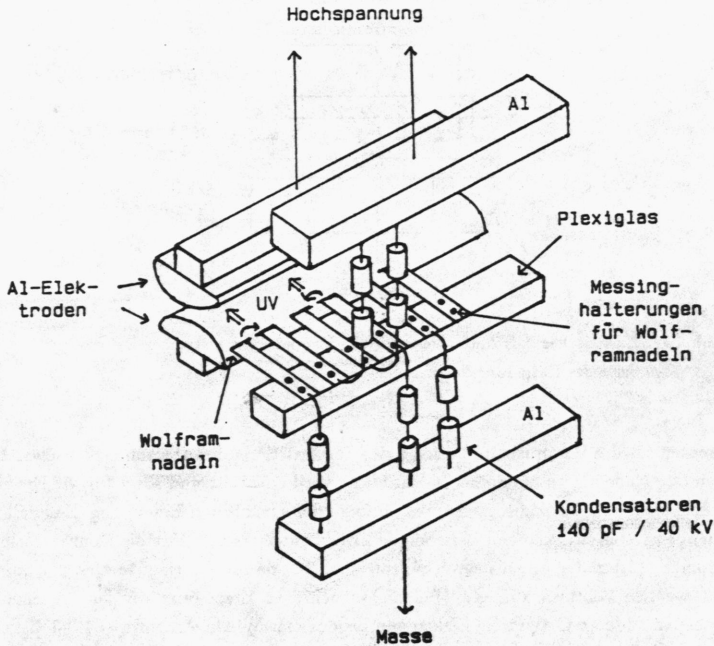


Abb. 9 Prinzipskizze für die Anordnung der Vorionisierungsfunknestrecken auf einer Seite des Entladungsvolumens.

den Funkenstrecken zu vermeiden.

Die Vorteile der Parallelschaltung der Funkenstrecken sind:

- 1) Die Induktivität des Vorionisierungsschaltkreises wird reduziert; man erhält kürzere und stärkere Strompulse über die Funkenstrecken und damit kürzere und intensivere UV-Strahlungspulse als bei einer seriellen Schaltung der Funkenstrecken.
- 2) Die Vorionisierung erfolgt nahezu synchron an allen Orten innerhalb des Entladungsvolumens.

Der letzte Punkt ist die Grundvoraussetzung dafür, daß der Vorionisierungskreis direkt parallel zum Hauptentladungskreis über die Elektroden gelegt werden kann. Die Ursache ist folgende: Noch während sich die Hochspannung (nach dem Zünden des im folgenden Abschnitt beschriebenen Marxgenerators) an den Hauptelektroden aufbaut, beginnen die Fun-

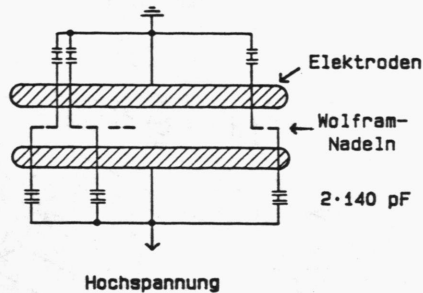


Abb. 10 Elektrisches Anschlußschema für die Vorionisierungsstrecken auf einer Seite des Entladungsvolumens.

kenstrecken für die Vorionisierung infolge des dort existierenden starken elektrischen Feldes (zugespitze Nadeln!) zu zünden. UV-Strahlung wird emittiert und sorgt für die Erzeugung freier Elektronen im Entladungsvolumen. Infolge der synchronen Erzeugung dieser Elektronen an allen Orten im Volumen zwischen den Elektroden wächst dort die Konzentration der Elektronen gleichmäßig an, bis sie genügend hoch ist, damit die Hauptentladung homogen über das ganze Volumen zündet. (Bei nichtsynchrone Erzeugung würden an einem Ort früher als an anderen Orten freie Elektronen produziert. Wie in Abschnitt I.1.2.3 diskutiert, wären Bogenentladungen die Folge.)

Dadurch, daß der gesamte Vorionisierungskreis parallel zum Hauptentladungskreis liegt, läßt er sich komplett innerhalb der Hochdruckkammer (vgl. Abschnitt I.1.3.4) unterbringen. Man erreicht damit eine sehr kompakte Laserkonstruktion, die auch nur eine Hochspannungsversorgung benötigt. Die hier beschriebene Anordnung stellt damit eine entscheidende Vereinfachung gegenüber den bisher veröffentlichten Vorionisierungsanordnungen dar [16,29,63]. Dort war nämlich die Reihe der Funkenstrecken auf jeder Seite des Entladungsvolumens generell seriell geschaltet. Die Folge war neben einer relativ hohen Induktivität des Vorionisierungskreises eine nichtsynchrone Vorionisierung der einzelnen Teile des Gasvolumens zwischen den Elektroden infolge der Laufzeit der Entladung längs der Funkenstrecken. Die Zeit, bis eine gleichmäßige Konzentration freier Elektronen aufgebaut war, mußte daher zur Vermeidung von Bogenentladungen abgewartet werden. Aus diesem Grund wurden 2 getrennte Hochspannungsversorgungs-kreise für die Vorionisierung und die Hauptentladung benutzt, die zeitlich gegeneinander verzögert (typisch einige μsec [16,63]) gezündet wurden. Der experimentelle Aufwand war also bei weitem größer als bei dem hier beschriebenen Laser.

1.3.3 Hochspannungsversorgung

Die Hochspannungspulse wurden von einem zweistufigen Marxgenerator geliefert, dessen elektrische Schaltung (Spannungsverdopplerschaltung nach Marx-Goodlet) in Abb.11 gezeigt ist. Dabei sei zunächst nur eine Entladungsstrecke im Laser betrachtet. Das Grundprinzip des Marxgenerators ist leicht erklärt. Es besteht in der parallelen Aufladung der beiden Kondensatoren C über die Widerstände R auf die Spannung U , die sich dann, durch Funkenstrecken seriell geschaltet, über die Elektroden des Lasers entladen. Infolge der seriellen Schaltung der Kondensatoren liegt damit zu Beginn des Entladungsprozesses eine Spannung vom Betrag $2U$ an den Laserelektroden an.

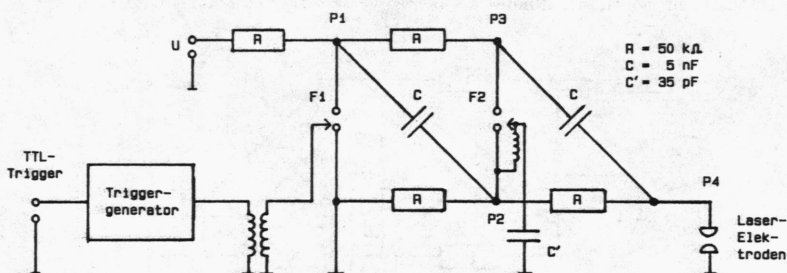


Abb.11 Schaltung des zweistufigen Marxgenerators zur Versorgung einer Entladungsstrecke.

Dieser Vorgang sei nun etwas genauer betrachtet: Vor der Zündung der Funkenstrecke F1, die zum Zeitpunkt t_0 erfolgen möge, hat die Spannung an den Punkten P1 und P3 den Wert U ; die Punkte P2 und P4 liegen auf Massepotential. F1 wird mit Hilfe eines Triggerelements (TR 050 der Fa. Tachisto Inc.) gezündet, der auf einen TTL-Triggerpuls hin einen -200 V-Puls mit einer Energie von 50 mJ und einer Anstiegszeit von 250 nsec liefert. Dieser Puls wird über Luftspulen auf -20 kV hochtransformiert und über eine Zündkerze an die Hilfselektrode der Funkenstrecke gegeben. Letztere schaltet daraufhin durch und legt den Punkt P1 schlagartig auf Masse. Die Spannung am Punkt P2 beträgt daher zum Zeitpunkt t_0 $-U$ (man vergleiche die Reaktion eines RC-Hochpasses auf das Anlegen einer Rechteckspannung /65/). Zwischen den Elektroden der Funkenstrecke F2 liegt somit eine Spannungsdifferenz vom Betrag $2U$. Die Durchbruchspannung U_D dieser Funkenstrecke, die durch den Gasdruck

in ihrem Inneren geregelt werden kann, ist so eingestellt, daß $U < U_D < 2U$ gilt. F2 schaltet also ebenfalls zum Zeitpunkt t_0 , so daß an der hochspannungsseitigen Elektrode des Lasers (Punkt P4) die Spannung $-2U$ anliegt. (Sie wird daher als Kathode bezeichnet.) Um ein zuverlässiges Zünden der Funkenstrecke F2 zu garantieren, war deren Hilfselektrode über einen Kondensator C' (realisiert aus 4 seriell geschalteten Keramik Kondensatoren von je 140 pF) mit Masse und über eine Spule (~ 20 Windungen, \varnothing 1.5 cm, Länge 10 cm) mit der Kathode der Funkenstrecke verbunden. Diese Schaltung hält zum Zeitpunkt t_0 die Hilfselektrode noch kurzzeitig auf Massepotential, wenn an der Kathode von F2 bereits die Spannung $-U$ anliegt. Die daraufhin erfolgende Entladung zwischen Hilfselektrode und Kathode garantiert das Durchschalten von F2.

Die Erweiterung des Marxgenerators für die Versorgung von 2 Entladungsstrecken zeigt Abb. 12. Die Verwendung gemeinsamer Funkenstrecken für die beiden Entladekreise gewährleistet ein synchrones Zünden der beiden Entladungsstrecken des Lasers.

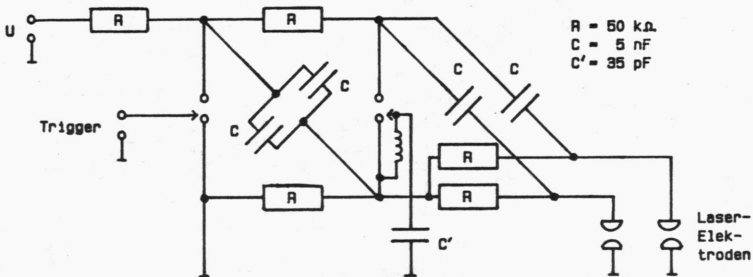


Abb. 12 Erweiterung des Marxgenerators zur Versorgung von 2 Entladungsstrecken.

Die Kapazität C der Kondensatoren betrug 5 nF. Sie wurde durch 5 parallel geschaltete 1 nF-Kondensatoren realisiert, die ihrerseits aus 2 seriell geschalteten 2 nF-Keramik Kondensatoren mit der Durchbruchspannung 40 kV bestanden. Durch die serielle Schaltung war es möglich, den Marxgenerator mit einer Spannung von 80 kV zu laden, ohne daß die Durchbruchspannung der Kondensatoren überschritten wurde. Die Widerstände R waren aus je 5 seriell geschalteten Hochspannungskohlewiderständen mit dem Wert 10 k Ω (Durchbruchspannung ebenfalls 40 kV) aufgebaut. Die Funkenstrecken (SG 501 der Fa. Tachisto Inc.) waren mit N_2 -Gas unter einem Druck von 3.5 atm gefüllt. Die Durchbruchspannung U_D betrug in diesem Fall etwa 75 kV; für ein ordnungsgemäßes Schalten war eine Mindestspannung von

30 kV zwischen den Elektroden der Funkenstrecke erforderlich. Während des Betriebs wurde der Stickstoff in den Funkenstrecken mit einer Rate von etwa 50 l/h ausgetauscht, um die Entladungsprodukte (z.B. N_2^+ -Ionen) zu entfernen.

Als Hochspannungsquelle wurde ein DC-Hochspannungsnetzgerät 0-100 kV/0-5 mA (HNC 100000 der Fa. Heinzinger) verwendet. Der maximale Strom von 5 mA, den dieses Netzteil liefern konnte, war der begrenzende Faktor für die Wiederholrate der Laserentladungen. Bei der Gesamtkapazität von 20 nF des Marxgenerators und einer Ladespannung von 60 kV dauert der Ladevorgang bei einem maximalen Strom von 5 mA etwa $\frac{1}{4}$ sec. Die maximale Wiederholrate beträgt somit 4 Hz. In der Praxis wurden jedoch maximal 2 Hz erreicht. Typische Wiederholfrequenzen lagen allerdings bei 0.5 Hz, um die Belastung der einzelnen Komponenten geringer zu halten und ein Aufheizen des Gasgemisches zu vermeiden.

Der gesamte Marxgenerator war in einem Edelstahlbehälter untergebracht, der zur elektrischen Isolation mit N_2 -Gas unter einem Druck von 6 atm gefüllt war. Bei der elektrischen Verschaltung wurde darauf geachtet, möglichst breite Bandleiter zu verwenden, um die Induktivität des Schaltkreises so gering wie möglich zu halten. Aus dem gleichen Grund bestanden die Verbindungen vom Marxgenerator zu den beiden Entladungsstrecken des Lasers aus je 4 parallel geschalteten Kabeln (vgl. Abb. 14). Als Kabel wurden dabei Koaxialkabel des Typs RG-213/U (garantierte Spannungsfestigkeit 70 kV) verwendet. Die Masseverbindung zwischen Marxgenerator und Druckkammer des Lasers erfolgte über Cu-Gewebebänder, die über die Koaxialkabel gezogen waren.

Zum Schluß sei noch bemerkt, daß der tatsächliche Betrag der Ausgangsspannung des Marxgenerators sogar höher als $2U$ ist. Die Ursache liegt in der Induktivität L des Marxgenerators bzw. der Verbindungskabel und der Kapazität C' dieser Kabel sowie der Laserentladungsstrecken selbst. Man erhält das vereinfachte Ersatzschaltbild in Abb. 13. C ist die Kapazität des Marxgenerators nach dem Zünden. Die Spannung U_K an den Kathoden des Lasers nach dem Zünden des Marxgenerators zum Zeitpunkt $t = 0$, der mit der Spannung U geladen wird, bis zum Einsetzen der Entladung läßt sich nach folgender Formel berechnen [57]:

$$U_K(t) = \frac{2UC}{C + C'} \left(1 - \cos\left(\sqrt{\frac{C + C'}{LC C'}} t\right) \right) \quad (18)$$

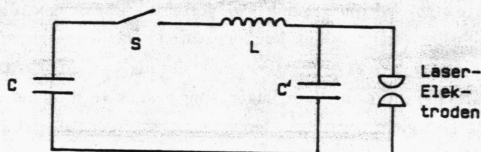


Abb. 13
Ersatzschaltbild der
Entladungsstrecke.

Die maximale Spannung beträgt somit $2U \frac{2C}{C+C'}$. C' läßt sich bei der verwendeten Anordnung zu etwa 2 nF abschätzen, so daß die maximale Spannung an den Kathoden des Lasers den Wert $2.8 U$ annimmt. Typische Versorgungsspannungen des Marxgenerators waren 60 kV (siehe Abschnitt I.1.4). Die Spannung an den Kathoden erreichte damit 170 kV und war nach der Diskussion in Abschnitt I.1.2.2 groß genug, um eine selbständige Gasentladung für ein Gasgemisch $\text{CO}_2:\text{N}_2:\text{He}=3:1:60$ bei einem Druck vom 20 atm aufrechtzuerhalten. Die im Marxgenerator gespeicherte Energie betrug in diesem Fall 36 J .

1.3.4. Mechanischer Aufbau

Eine Skizze des mechanischen Aufbaus zeigt Abb. 14. Die Druckkammer bestand aus 2 Edelstahlzylindern mit einem Durchmesser von 40 cm , einer Länge von 60 cm und einer Wandstärke von 9.5 mm , an deren Enden jeweils Edelstahlflansche aufgeschweißt waren. Die beiden miteinander verschraubten Zylinder waren an ihren offenen Enden mit Aluminiumplatten der Dicke 5 cm verschlossen. Alle Dichtungen erfolgten mit O-Ringen. Die so gebildete Kammer war druckfest bis über 25 atm .

Der optische Zugang zur Druckkammer wurde durch NaCl-Brewsterfenster ermöglicht. Diese Fenster aus NaCl-Einkristallen (Fa. Korth) mit einem Durchmesser von 35 mm und einer Dicke von 10 mm waren mit einem handelsüblichen Plastikkleber auf einen Aluminiumeinsatz aufgeklebt, der zur Strahlführung eine Bohrung von 1 cm Durchmesser besaß und der auswechselbar von außen auf die seitlichen Aluminiumplatten geschraubt war. Auf diese Weise war ein problemloser Fensterwechsel möglich, ohne daß die Druckkammer zerlegt werden mußte. Die Fenster waren gegen Feuchtigkeit von außen durch mit Kieselgel gefüllte Aufsätze geschützt.

Für die Hochspannungsdurchführungen (in Abb. 14 sind der Übersichtlichkeit halber nur eine statt der tatsächlich jeweils vier pro Entladungsstrecke gezeichnet) wurden verschiedene Plastikmaterialien getestet: Nylon, Delrin und rotes PVC. Letzteres erwies sich als den Beanspruchungen am ehesten gewachsen. Bei den anderen Materialien traten infolge der mit den Entladungen verbundenen mechanischen Verspannungen im Laufe der Zeit Materialermüdungen auf, so daß Überschlänge der Hochspannung an die Laserwand möglich waren. Dennoch traten diese Probleme, wenn auch weit weniger häufig, auch bei rotem PVC auf.

Die Elektroden und die zugehörigen Vorionisierungsstrecken einer jeden Entladungsstrecke waren inklusive der Kopplungskondensatoren auf einen eigenen Träger montiert, der ebenfalls komplett austauschbar war. Der elektrische Kontakt zwischen den Kathoden und den Hochspannungszuführungen erfolgte über jeweils 16 gehärtete Blattfedern aus Stahl (siehe Abb. 14). Das gleiche galt für den Kontakt zwischen den Anoden und der Laserwand, die

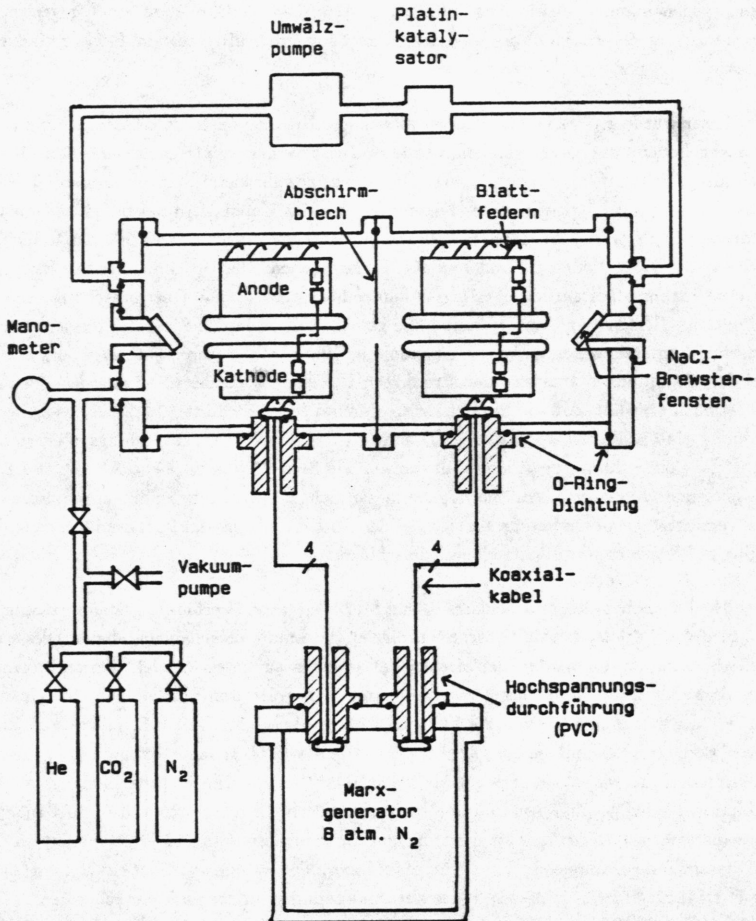


Abb.14 Mechanischer Aufbau (nicht maßstabsgerecht). Der Übersichtlichkeit halber ist nur eine Hochspannungsdurchführung in jeder Laserdruckkammer statt der tatsächlich jeweils vier eingezeichnet.

Massepotential besitzt. Die beiden Entladungsstrecken wurden durch ein auf Masse liegendes Blech gegeneinander abgeschirmt; ohne diese Abschirmung kam es zu einer gegenseitigen Beeinflussung der Entladungsstrecken, was zu häufigem Auftreten von Bogenentladungen führte.

Der Laser wurde im 'sealed off'-Betrieb verwendet. Um eine bestimmte Gasmischung bei einem gegebenen Druck p einzustellen, wurde zunächst die gesamte Druckkammer mittels einer Vakuumpumpe auf 0.1 Torr evakuiert. Damit war gewährleistet, daß keine unerwünschten Gasreste, wie z.B. O_2 mehr in der Kammer vorhanden waren. Um beispielsweise ein Gasgemisch $CO_2:N_2:He = 3:1:60$ bei $p = 20$ atm zu erzeugen, wurden daraufhin der Reihe nach 900 mbar CO_2 , 300 mbar N_2 und 18.8 atm He zugegeben. Der Druck konnte mit Hilfe eines Feinmeßmanometers auf 50 mbar genau kontrolliert werden. Die Durchmischung der Gase erfolgte mit Hilfe einer Umwälzpumpe, die gemäß Abb. 14 in das System integriert war. In diesem Pumpkreis waren gleichzeitig keramische Platinkatalysatoren enthalten, um aus den bei den Entladungen entstehenden Produkten CO und O_2 wieder CO_2 zurückzugewinnen. Es sei noch erwähnt, daß die Vorionisierung durch die in Abschnitt I.1.3.2 beschriebene Anordnung eine genügend hohe Konzentration freier Elektronen im Entladungsvolumen auch ohne die Beimischung organischer Substanzen wie Trimethylamin (siehe Abschnitt I.1.2.3) ins Gasgemisch erzeugte. Auf die Zugabe dieser nicht ungefährlichen Substanzen wurde daher verzichtet, insbesondere da sie sich bei der Entladung zersetzen, polymerisieren und als Belag auf Fenstern und Elektroden niederschlagen.

Der Hochdruckbehälter war einschließlich Marxgenerator, Verbindungskabel, Hochspannungsversorgung etc. in einer Abschirmkabine aus Aluminiumblechen aufgebaut. Dies erwies sich als nötig, da die bei den elektrischen Entladungen auftretenden elektrischen Störungen die Verwendung empfindlicher Detektoren und Meßgeräte unmöglich machten. (In einer Entfernung von 3 m betrug die Störungen ohne Abschirmkabine z.B. einige Volt an einem Oszillographen mit einem Eingangswiderstand von 50 Ω , aber offenem Eingang; mit Abschirmkabine waren sie geringer als 0.5 mV!). Der optische Zugang in die Abschirmkabine war möglich über Aluminiumröhrchen mit 2 cm Durchmesser und 9 cm Länge, die eine ausreichende Dämpfung für elektromagnetische Strahlung mit einer Frequenz ≤ 1 GHz aufwiesen. Die gesamte elektrische Anlage im Innern der Kabine war über ein Netzfilter (Siemens B84261 A13 B11), dessen Wechselstromdämpfung im Frequenzbereich von 100 KHz bis 30 GHz 100 dB betrug, vom externen Leitungsnetz entkoppelt. Die innerhalb der Kabine nötigen Steuersignale, wie z.B. Triggerpuls zum Zünden des Marxgenerators, Steuerspannung für das Hochspannungsnetzteil etc. wurden über Lichtleiter übertragen. Analoge Signale wurden dabei mit Hilfe von Spannungs-Frequenz-Wandlern in Rechteckpulse umgewandelt, deren Frequenz zu dem analogen Spannungssignal proportional war. Zur Demodulation wurden 'phase-locked-loop'-Schaltkreise verwendet.

1.3.5 Resonatorkonfiguration

Von den vielen getesteten Resonatorkonfigurationen sei hier diejenige diskutiert, die letztendlich zum Pumpen des CH_3F -Raman-Lasers verwendet wurde. Sie ist in Abb. 15 skizziert. Die wesentlichen Elemente des Resonators, der eine Gesamtlänge von 3.3 m besaß, waren:

- Ein konkaver Ge-Spiegel mit einem Krümmungsradius $R = 25 \text{ m}$ und einem Reflexionsvermögen von 70% für Strahlung mit der Wellenlänge $10.6 \mu\text{m}$, der als Auskoppelspiegel benutzt wurde;
- ein Strahlaufweitungsteleskop 1:3, das von einem konvexen bzw. konkaven Molybdänspiegel mit Krümmungsradius $R = 0.3 \text{ m}$ bzw. $R = 1 \text{ m}$ im Abstand von 35 cm gebildet wurde;
- ein Gitter (150 Striche/mm, Größe $5 \times 5 \text{ cm}^2$) mit einer Blazewellenlänge von $10.6 \mu\text{m}$, das als frequenzselektiver hochreflektierender Spiegel wirkt (das Reflexionsvermögen beträgt nach Herstellerangaben für Strahlung, die senkrecht zu den Gitterstrichen polarisiert ist, mehr als 95% im Wellenlängenbereich von $9 \mu\text{m}$ bis $11 \mu\text{m}$).

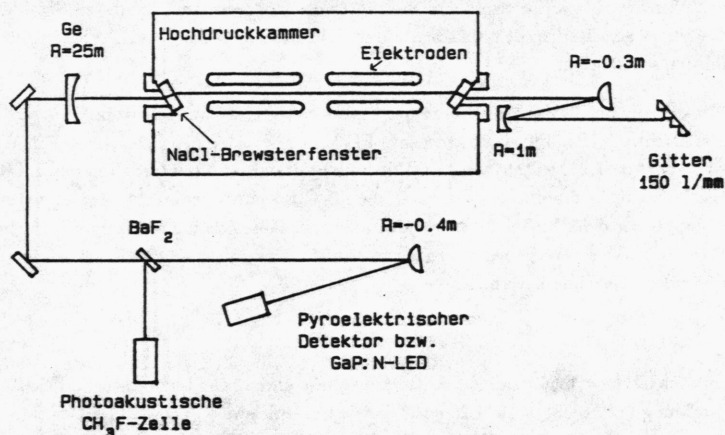


Abb. 15 Resonatorkonfiguration (alle verwendeten Spiegel waren aus Molybdän).

Das Strahlaufweitungsteleskop dient zum einen dazu, die Intensität der Laserstrahlung auf der Gitteroberfläche zu reduzieren, um eine Schädigung derselben durch Plasmadurchbrüche

zu vermeiden, die bei einer Oxidation der Oberfläche oder infolge von darauf befindlichen Staubpartikeln bei sehr hohen elektrischen Feldstärken auftreten. Zum anderen bedeutet eine volle Ausleuchtung des Gitters die bestmögliche spektrale Auflösung und somit die geringste spektrale Breite der emittierten Strahlung. Das Gitter war mit vertikalen Strichen und vertikaler Drehachse auf einem Drehtisch montiert. Die emittierte Laserstrahlung war somit horizontal polarisiert. Dementsprechend waren die NaCl-Brewsterfenster orientiert.

Der Durchmesser $2w_0$ der TEM₀₀-Resonatormode (e^{-1} -Breite der Feldstärke) würde für die Resonatorkonfiguration in Abb.15 theoretisch 6 mm am Auskoppelspiegel betragen. Da der Elektrodenabstand der beiden Entladungsstrecken jedoch nur 1 cm beträgt, ist sofort einsichtig, daß die Ausbildung einer reinen TEM₀₀-Mode nicht möglich ist. Die metallischen Elektroden wirken als Wellenleiter. Um dennoch eine symmetrische Mode im Resonator zu erhalten, waren die Elektrodenpaare der beiden Entladungsstrecken um 90° gegeneinander verdreht. Die Querschnittsverteilung der Strahlung außerhalb des Resonators konnte dann annähernd mit einer Gaußschen Verteilung, wie sie eine reine TEM₀₀-Mode liefern würde, beschrieben werden. Ein besonders kritischer Parameter war dabei der Abstand der beiden Teleskopspiegel. Dieser bestimmt nämlich den Krümmungsradius der als sphärischer Spiegel wirkenden Kombination von Teleskop und Gitter. Insbesondere mußte auf einen möglichst kleinen Winkel - im vorliegenden Fall betrug er etwa 5° - in dem gefalteten Strahlengang des Teleskops geachtet werden, da sonst durch die Komausbildung starke Verzerrungen des Strahlprofils auftraten.

Die Einstellung der Emissionsfrequenz des Lasers erfolgte in zwei Schritten:

- 1) Der Drehtisch (Fa. Physik Instrumente PI) konnte durch einen Schrittmotor in Winkelschritten von 7.2" gedreht werden. Wird die Position der 0.ten Ordnung des in Littrowkonfiguration angeordneten Gitters, an der die Gitterebene senkrecht zur Resonatorachse steht, durch den Winkel $\theta = 0$ gekennzeichnet, läßt sich die Winkelposition, bei der Strahlung einer vorgegebenen Frequenz ν in sich selbst zurückreflektiert wird, anhand folgender Gleichung berechnen:

$$2a\nu \sin \theta = c \quad (19)$$

Hierbei ist $a = 6.666 \mu\text{m}$ die Gitterkonstante und c die Lichtgeschwindigkeit. Für Strahlung der Frequenz 1074.36 cm^{-1} beträgt beispielsweise $\theta = 44.33^\circ$. Das sind 22165 Winkelschritte. Statt der Position der 0.ten Ordnung läßt sich auch die Gitterstellung als Referenzposition verwenden, bei der ein He-Ne-Laserstrahl in 16.ter Ordnung in sich selbst zurückreflektiert wird. Diese Position entspricht einer Frequenz 987.65 cm^{-1} .

- 2) Da die Gitterkonstante a nur auf $\pm 0.001 \mu\text{m}$ genau bekannt war, konnte die Winkelposition des Gitters für eine gewünschte Emissionsfrequenz jedoch nicht exakt bestimmt werden. Beispielsweise kann die gesuchte Gitterposition für die Frequenz 1074.36 cm^{-1} nur auf ± 4 Winkelschritte genau angegeben werden. Um die Emissionsfrequenz des

Hochdruck-CO₂-Lasers und die Gitterstellung einander exakt zuordnen zu können, wurde ein Teil der emittierten Laserstrahlung mittels einer BaF₂-Scheibe in eine Gaszelle eingekoppelt, die CH₃F-Gas bei einem Druck von 30 Torr und ein Mikrophon enthielt (siehe Abb. 15). Absorption der Laserstrahlung im CH₃F-Gas führt zu einer Erwärmung des Gases und damit zu Druckänderungen in der Zelle, die vom Mikrophon registriert werden (photoakustischer Effekt /66/). Die Frequenzen der CH₃F-Absorptionslinien sind sehr genau bekannt (vgl. Abschnitt I.2.1.1 bzw. /63,67/) und überdecken mit einem typischen Abstand von 1-2 cm⁻¹ einen Großteil des Emissionsbereichs des CO₂-Lasers. Sie wurden daher jeweils als Referenzfrequenzen zur exakten Eichung der Gitterstellung benutzt. Die Einstellung von Emissionsfrequenzen, die zwischen den CH₃F-Absorptionslinien lagen, erfolgte durch lineare Interpolation. Eine relative Frequenzänderung der emittierten Strahlung war dabei in Schritten von 0.04 cm⁻¹ bzw. 1.2 GHz möglich. Dies folgt aus Gl.(19) unter Berücksichtigung der Schrittweite von 7.2'' des schrittmotorgetriebenen Drehtisches.

Die Messung der Pulsenergie der CO₂-Laserstrahlung erfolgte mit Hilfe eines pyroelektrischen Detektors (Gentec ED500, Empfindlichkeit $2.4 \frac{\text{V}}{\text{J}}$, Detektorfläche 5x5 cm²). Um eine Zerstörung des Detektors durch zu hohe Energiedichten zu vermeiden, wurde der Laserstrahl durch einen konvexen Molybdänspiegel aufgeweitet. Zur Beobachtung der zeitlichen Struktur der Laserpulse wurde eine als Photoleitungsdetektor wirkende, auf 4.2 K gekühlte GaP:N-Leuchtdiode verwendet, die eine Anstiegszeit von weniger als 0.5 nsec besitzt und die für Strahlung der Wellenlänge 10.6 μm bei einer angelegten Spannung von 80 V eine Empfindlichkeit von $0.3 \frac{\text{V}}{\text{W}}$ bei einem Lastwiderstand von 50 Ω besitzt /68,69/. Die von der Diode gelieferten Signale wurden mit einem schnellen Transientenrekorder (Biomation 6500, Abtastrate 500 MHz, Eingangsbandbreite 125 MHz) aufgezeichnet und mit Hilfe eines Z80-ECB-Bus-Rechners verarbeitet. Dieser Rechner übernahm dabei auch noch die komplette Steuerung des Schrittmotors für den Gitterdrehtisch (vgl. Abb. 34), die Kontrolle der Hochspannungsversorgung des Marxgenerators sowie dessen Triggerrung.

1.4 Eigenschaften des Hochdruck-CO₂-Lasers

Die Eigenschaften des im vorhergehenden Abschnitt beschriebenen Hochdruck-CO₂-Lasers wurden bei verschiedenen Resonatorkonfigurationen sowie bei Verwendung von einer bzw. zwei Entladungsstrecken untersucht. In diesem Abschnitt werden die wesentlichen Ergebnisse zusammengefaßt. Insbesondere wird dabei zum ersten Mal die vollständige Emissionscharakteristik eines bei einem Druck von 20 atm betriebenen CO₂-Lasers vorgestellt.

1.4.1 Abstimmcharakteristik

Abb. 16 a zeigt die Ausgangspulsenergie des Lasers als Funktion der Emissionsfrequenz für einen Betriebsdruck von 20 atm bei einer Gasmischung $\text{CO}_2:\text{N}_2:\text{He}$ 3:1:60 und bei Verwendung von nur einer Entladungsstrecke. Die Emissionsfrequenz wurde dabei in Schritten von 0.12 cm^{-1} variiert, die Pulsenergie über 3 aufeinanderfolgende Pulse gemittelt. Das aktive Gasvolumen betrug $1 \times 1 \times 40 \text{ cm}^3$ (eine Entladungsstrecke!); die in dieses Volumen injizierte Energie war 21 J, was einer Ladespannung $U = 65 \text{ kV}$ für den Marxgenerator entsprach. Als Entladungsenergie-dichte erhält man daher $26 \frac{\text{J}}{\text{l atm}}$. Die Resonatorkonfiguration war die in Abb. 15 gezeigte Konfiguration; die gesamte Resonatorlänge betrug jedoch nur 2.2 m infolge der fehlenden zweiten Entladungsstrecke.

Abb. 16 a demonstriert die Möglichkeit zur kontinuierlichen Abstimmung der Laseremissionsfrequenzen in allen 4 Emissionszweigen des Lasers. Der gesamte Abstimmbereich betrug 44 cm^{-1} ; die Abstimmbereiche in den einzelnen Zweigen sind in der 1. Spalte von Tab. 4 zusammengefaßt.

Zweig	Abstimmbereich	
	Entladungsenergie-dichte $26 \frac{\text{J}}{\text{l atm}}$	Entladungsenergie-dichte $35 \frac{\text{J}}{\text{l atm}}$
10 R	$R(8)-R(26) 13 \text{ cm}^{-1}$	$R(6)-R(28) 15 \text{ cm}^{-1}$
10 P	$P(12)-P(26) 14 \text{ cm}^{-1}$	$P(10)-P(28) 18 \text{ cm}^{-1}$
9 R	$R(10)-R(26) 10 \text{ cm}^{-1}$	$R(8)-R(30) 15 \text{ cm}^{-1}$
9 P	$P(14)-P(20) 7 \text{ cm}^{-1}$	$P(12)-P(24) 13 \text{ cm}^{-1}$

Tab. 4 Abstimmbereiche in den einzelnen Emissionszweigen bei Verwendung von nur einer Entladungsstrecke.

Die Emissionscharakteristik des Lasers in den beiden *R*-Zweigen ist total glatt, während in den beiden *P*-Zweigen noch eine geringfügige Welligkeit vorhanden ist. Dies ist die Folge des größeren Abstands benachbarter Rotations-Schwingungs-Übergänge in den *P*-Zweigen (1.8 cm^{-1}) gegenüber dem in den *R*-Zweigen (1.3 cm^{-1}) - siehe Abschnitt I.1.1. Der verglichen mit den anderen Zweigen geringe Abstimmbereich des 9*P*-Zweiges und die dortige geringe Ausgangspulsenergie ($\sim 30 \text{ mJ}$) sind konsistent mit den von Taylor et al. /47/ durchgeführten Berechnungen der Kleinsignalverstärkung, die im 9*P*-Zweig eine um etwa 10% geringere Verstärkung voraussagen, als sie in den restlichen Zweigen existiert (Abb. 3). Daß im vorliegenden Fall die maximale Ausgangspulsenergie im 10*R*-Zweig erhalten wurde, die nach Taylor et al. im 9*R*-Zweig erhalten werden mußte, kann eine Folge davon sein,

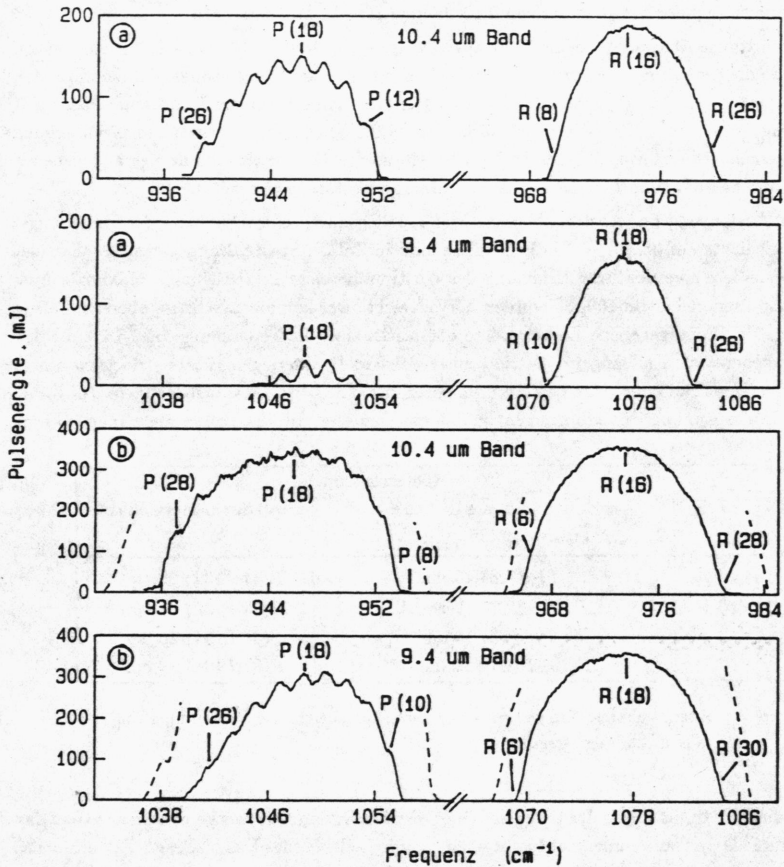


Abb. 16 Emissionscharakteristik des Hochdruck-CO₂-Lasers bei $p = 20$ atm

- a) bei einer Entladungsstrecke und einer Entladungsenergie-dichte $\varepsilon = 26 \frac{\text{J}}{\text{l atm}}$,
 b) bei zwei Entladungsstrecken und einer Entladungsenergie-dichte $\varepsilon = 21 \frac{\text{J}}{\text{l atm}}$ in den R-Zweigen bzw. $\varepsilon = 26 \frac{\text{J}}{\text{l atm}}$ in den P-Zweigen.
 Die gestrichelten Kurven zeigen die Ausweitung des Abstimm-bereichs bei einer Entladungsenergie-dichte $\varepsilon = 30 \frac{\text{J}}{\text{l atm}}$.

daß die Justierung des Resonators im 9R-Zweig nicht optimal war (man vergleiche auch die Ergebnisse bei der Verwendung von 2 Entladungsstrecken in Abb. 16 b).

Die maximale Ausgangspulsenergie im 10R-Zweig betrug 180 mJ. Da hierbei eine Energie von 21 J in die Entladung injiziert wurde, erhält man somit einen Energiekonversionskoeffizienten von etwa 1% im Zentrum des 10R-Zweiges. Der maximal erreichbare Konversionskoeffizient beträgt etwa 8%. Dies wurde in Abschnitt I.1.2.2 für die hier verwendete Gasmischung und einen $\frac{E}{N}$ -Wert von $2.5 \cdot 10^{-16} \text{ V cm}^2$ - ein solcher entspricht gerade der angegebenen Versorgungsspannung $U = 65 \text{ kV}$ für den Marxgenerator - abgeschätzt.

Eine Ausweitung des Abstimmbereichs war möglich durch eine Erhöhung der Entladungsenergiedichte auf $35 \frac{\text{J}}{\text{cm}^3 \text{ atm}}$. Die Ergebnisse sind in Tab. 4, Spalte 2 angegeben. Man findet insbesondere eine deutliche Erhöhung des Abstimmbereichs im 9P-Zweig. Maximale Ausgangspulsenergien von 150 mJ wurden dort erreicht. Der komplette Abstimmbereich betrug 62 cm^{-1} . Die angegebene Energiedichte entspricht einer Ladespannung von 75 kV für den Marxgenerator. Ein längerer Betrieb unter solchen Bedingungen führte jedoch zu einem starken Verschleiß der einzelnen Komponenten des Entladekreises, insbesondere der Kabel und Hochspannungsdurchführungen, so daß ein Dauerbetrieb des Lasers nicht möglich war.

Zweig	Abstimmbereich	
	Entladungsenergiedichte $21 \frac{\text{J}}{\text{cm}^3 \text{ atm}}$ (R-Zweig), $26 \frac{\text{J}}{\text{cm}^3 \text{ atm}}$ (P-Zweig)	Entladungsenergiedichte $30 \frac{\text{J}}{\text{cm}^3 \text{ atm}}$
10 R	$R(6) - R(28) 15 \text{ cm}^{-1}$	$R(4) - R(34) 20 \text{ cm}^{-1}$
10 P	$P(8) - P(28) 18 \text{ cm}^{-1}$	$P(6) - P(32) 24 \text{ cm}^{-1}$
9 R	$R(6) - R(30) 16 \text{ cm}^{-1}$	$R(4) - R(34) 20 \text{ cm}^{-1}$
9 P	$P(10) - P(26) 16 \text{ cm}^{-1}$	$P(8) - P(30) 22 \text{ cm}^{-1}$

Tab. 5 Abstimmbereiche in den einzelnen Emissionszweigen bei Verwendung von zwei Entladungsstrecken.

Aus diesem Grund wurde der Laser mit einer zweiten Entladungsstrecke versehen, so daß das gesamte aktive Gasvolumen auf $1 \times 1 \times 80 \text{ cm}^3$ verdoppelt wurde. Das in diesem Fall erhaltene Emissionsspektrum des Lasers ist in Abb. 16 b gezeigt. Die Entladungsenergiedichte betrug hierbei $21 \frac{\text{J}}{\text{cm}^3 \text{ atm}}$ in den R-Zweigen (dies entspricht einer Marxgeneratorversorgungsspannung $U = 57 \text{ kV}$) und $26 \frac{\text{J}}{\text{cm}^3 \text{ atm}}$ ($U = 63 \text{ kV}$) in den P-Zweigen. Der gesamte Abstimmbereich war 65 cm^{-1} , eine Aufschlüsselung in die einzelnen Bereiche ist in Tab. 5, Spalte 1 gegeben. Man erhält somit bereits bei typischen Marxgeneratorladespannungen um 60 kV, bei denen ein problemloser Laserbetrieb möglich war, den gleichen Abstimmbereich, wie er unter den oben angegebenen extremen Bedingungen im Falle einer Entladungsstrecke erhalten wurde. Bei der Aufnahme der Emissionscharakteristik in Abb. 16 b mußte darauf geachtet werden,

daß die Pulsenergie 300 mJ nicht überschritt. Denn höhere Pulsenergien führten zu einer Schädigung der NaCl-Brewsterfenster durch Plasmadurchbrüche an ihren Oberflächen. Die gestrichelten Linien in Abb. 16 b zeigen die Ausweitung des Abstimmereichs bei Erhöhung der Entladungsenergiedichte auf $30 \frac{\text{J}}{\text{l atm}}$ (siehe auch Tab. 5). Die entsprechende Versorgungsspannung $U = 68 \text{ kV}$ des Marxgenerators war die maximale Spannung, die noch einen Betrieb ohne größere Probleme zuließ.

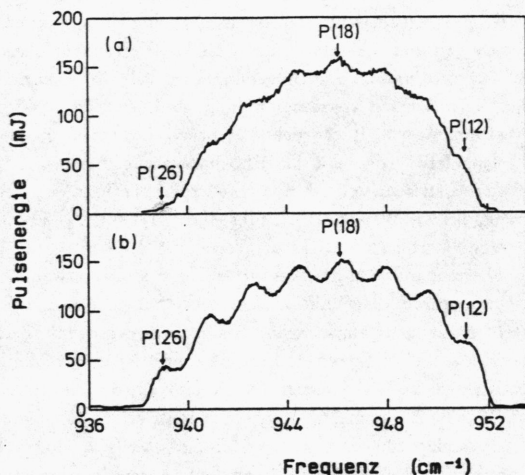


Abb. 17 Emissionscharakteristik des 10P-Zweiges bei einem Gasdruck von 25 atm (a) bzw. 20 atm (b).
(In beiden Fällen war die in die Entladung injizierte Energie 21 J.)

Um zu überprüfen, ob die Restwelligkeit der Emissionscharakteristik in den P-Zweigen durch Verwendung noch höherer Gasdrücke beseitigt werden könnte, wurde das Spektrum des 10P-Zweiges bei einem Druck von 25 atm (Gasgemisch $\text{CO}_2:\text{N}_2:\text{He} = 3:1:60$) aufgenommen. Abb. 17 a zeigt dieses Spektrum bei einer Entladungsenergiedichte von $21 \frac{\text{J}}{\text{l atm}}$. Hierbei war wiederum nur eine Entladungsstrecke im Laser vorhanden. In Abb. 17 b ist zum Vergleich noch einmal das Emissionsspektrum des 10P-Zweiges bei 20 atm aus Abb. 16 a dargestellt. Während die Modulationstiefe im Zentrum des Zweiges bei 20 atm noch etwa 15% beträgt, ist sie bei 25 atm auf die Hälfte reduziert. Da dies keine drastische Reduktion bedeutete und auch der Abstimmereich der gleiche blieb, wurde darauf verzichtet, den Laser bei Drücken über 20 atm zu betreiben.

Die Bandbreite der emittierten Strahlung (FWHM = volle Breite bei halber Höhe) betrug im Zentrum des 10R-Zweiges 0.16 cm^{-1} (4.8 GHz). Sie wurde mit Hilfe eines Infrarotspektrometers ermittelt, das ein Gitter mit 150 Strichen pro mm und 13 cm Breite und damit eine Auflösung von 0.05 cm^{-1} für Frequenzen in der Umgebung von 1000 cm^{-1} besaß. Mittels dieses Spektrometers wurde auch nachgewiesen, daß kein 'frequency pulling' die kontinuierliche Abstimmbarkeit einschränkt. Von 'frequency pulling' spricht man, wenn die Frequenzen der longitudinalen Lasermoden nicht mehr durch die Resonatoreigenfrequenzen $\nu_m = \frac{mc}{2L}$, $m \in \mathbb{N}$ (L ist die Resonatorlänge) gegeben sind, sondern in Richtung eines Maximums der Kleinsignalverstärkung verschoben sind. Ursache dafür ist die Frequenzabhängigkeit der komplexen Suszeptibilität des Lasermediums in der Umgebung eines solchen Maximums [39]. Mehr oder weniger starkes Auftreten von 'frequency pulling' wird in vielen Veröffentlichungen zur kontinuierlichen Abstimmung von Hochdruck-CO₂-Lasern berichtet [58,63], da dort stets Drücke $\leq 12 \text{ atm}$ verwendet wurden und daher noch deutliche Maxima und Minima in der Kleinsignalverstärkung existieren (vgl. Abb. 3). Ein experimentelles Beispiel für 'frequency pulling' im 10R-Zweig, das von Wan et al. [58] für einen 10 atm-CO₂-Laser erhalten wurde, ist in Abb. 18a gezeigt. Dort ist die spektrale Verteilung der emittierten Laserstrahlung, die mit Hilfe eines Thermopapiers in der Austrittsebene des oben erwähnten Infrarotspektrometers aufgezeichnet wurde, als Funktion der theoretischen, durch die Winkelposition des Gitters eingestellten Frequenz aufgetragen. Die endliche Ausdehnung der Färbung des Thermopapiers ist ein Maß für die spektrale Breite der emittierten Strahlung. Man erkennt deutlich, daß die emittierten Laserfrequenzen stets in der Nähe der Linienzentren liegen. Als Vergleich ist in Abb. 18b der Zusammenhang zwischen emittierter und eingestellter Frequenz für einen Teil des 10P-Zweiges für den hier vorgestellten Laser ($p = 20 \text{ atm}$) dargestellt, der sich als absolut linear erweist. Damit ist der eindeutige Beweis erbracht, daß der Laser kontinuierlich abstimmbar ist und daß die Emissionsfrequenz nur durch die Winkelposition des Gitters bestimmt wird. Dies ist die Konsequenz des nahezu glatten Profils der Kleinsignalverstärkung bei einem Gasdruck von 20 atm und zeigt, daß der Betrieb kontinuierlich abstimmbarer Hochdruck-CO₂-Laser erst bei einem solchen Druck sinnvoll wird.

1.4.2 Zeitliche Struktur der Laserpulse

Abb. 19 zeigt typische Ausgangspulse des Lasers bei einem Betrieb weit oberhalb der Laserschwelle (a) bzw. knapp an der Schwelle (b). Die Emissionsfrequenz lag für Puls (a) im Zentrum des 10 R-Zweiges (973.5 cm^{-1}), für Puls (b) in der Flanke (968.3 cm^{-1}). Die experimentellen Bedingungen waren die gleichen wie bei der Aufnahme der Emissionscharakteristik in Abb. 16a (nur eine Entladungsstrecke, $p = 20 \text{ atm}$, Gasgemisch CO₂:N₂:He = 3:1:60, Entladungsenergiegedichte $26 \frac{\text{J}}{\text{atm}}$). Die Pulsenergie betrug 180 mJ im Falle von Puls (a) und etwa 60 mJ für Puls (b). Das Einsetzen der elektrischen Störung in Abb. 19 (Insets) kennzeichnet

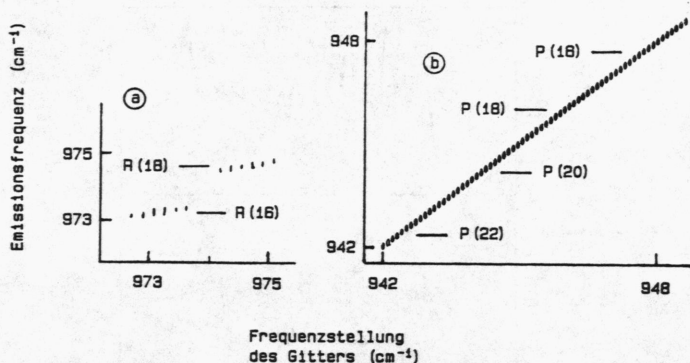


Abb. 18 Zusammenhang zwischen der Emissionsfrequenz der Laserstrahlung und der theoretischen, durch die Winkelposition des Gitters bestimmten Frequenz

- a) bei $p = 10$ atm nach /58/,
- b) bei $p = 20$ atm für den hier vorgestellten Laser.

dabei den Beginn der elektrischen Entladung (die Abschirmkabine war bei diesen Messungen nicht installiert!). Abb. 19 gestattet daher eine direkte Bestimmung der Aufbauzeit t_{auf} der Laserpulse, d.h. der Zeit, bis die Pulse nach Zünden der Entladung ihr Maximum erreichen. Man erhält für Puls (a) eine Aufbauzeit von 450 nsec und eine Pulsdauer von 130 nsec. Letztere wird bestimmt durch den Abbau der Überbesetzung durch die stimulierte Emission. Die Aufbauzeit für Puls (b) betrug 700 nsec infolge der geringeren Kleinsignalverstärkung in der Flanke des 10R-Zweiges, was auch der Grund für die längere Pulsbreite von 190 nsec ist. Zusätzlich ist zu bemerken, daß im Falle von Puls (b) die in Abschnitt I.1.1.3 erwähnte Relaxationszeit von 700 nsec für den V-T-Transfer der ν_3 -Normalschwingung bereits eine Rolle spielt.

Längere Aufbauzeiten und Pulsbreiten als bei Puls (b) können nicht mehr auftreten, da typisch 1 μ sec nach der Entladung Schockwellen auftreten, die durch die Vorionisierungs-entladungen bzw. durch die Hauptentladung selbst verursacht werden. Sie führen zu einem starken Anstieg der Resonatorverluste und machen so eine Lasertätigkeit unmöglich /47/.

Mit Hilfe der Aufbauzeit und Energie der Laserpulse läßt sich eine einfache Abschätzung der Kleinsignalverstärkung durchführen /62/. Es sei dazu angenommen, daß die Verstärkung gleichzeitig für alle longitudinalen Moden, die innerhalb der spektralen Breite $\Delta\nu_L$ der emit-

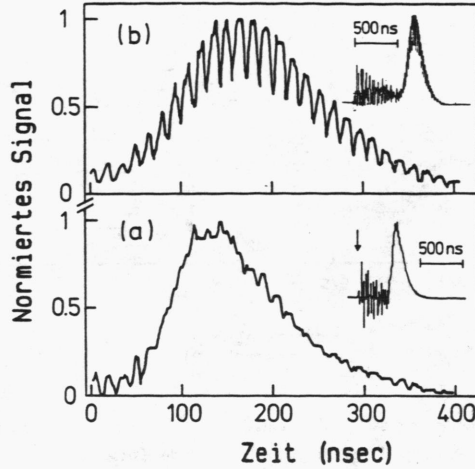


Abb. 19 Zeitliche Struktur der Laserpulse für einen Betrieb weit über der Laserschwelle (a) bzw. knapp an der Schwelle (b). Die Insets zeigen den Aufbauprozess der Laserpulse.

tierten Strahlung enthalten sind - ihre Gesamtzahl sei m -, zum Zeitpunkt der elektrischen Entladung durch spontane Emission eines Photons beginnt und nach der Aufbauzeit t_{auf} endet. Für die Zahl Z der Photonen, die in einer Mode nach der Zeit t_{auf} enthalten sind, gilt

$$Z = \exp \left((2l\alpha(\nu_L) - \beta) \cdot \frac{t_{auf}}{t_{um}} \right) \quad (20)$$

Hierbei ist l die Länge des aktiven Mediums und $\alpha(\nu_L)$ ist die Kleinsignalverstärkung bei der Emissionsfrequenz ν_L (sie kann für alle Moden als konstant angenommen werden, da sich die Verstärkung innerhalb der geringen Bandbreite $\Delta\nu_L$ nicht ändert). β bezeichne den Verlust pro Umlauf eines Photons im Resonator und t_{um} die für den Umlauf benötigte Zeit. Für die gesamte Energie des Laserpulses gilt dann

$$E = h\nu_L m Z = h\nu_L m \exp \left((2l\alpha(\nu_L) - \beta) \cdot \frac{t_{auf}}{t_{um}} \right) \quad (21)$$

Aus der Resonatorlänge von $L = 2.2$ m errechnet man eine Umlaufzeit $t_{um} = \frac{2L}{c} = 14.7$ nsec. Der Frequenzabstand der longitudinalen Moden beträgt daher 68 MHz. Innerhalb der Bandbreite $\Delta\nu_L = 4.8$ GHz sind damit $m = 70$ longitudinale Moden enthalten. Werden die

Resonatorverluste nur durch den Auskoppelspiegel mit Reflexionsvermögen 70% bestimmt, gilt $\beta = 0.3$. Im Falle des Pulses (a) ($E = 180 \text{ mJ}$, $\nu_L = 973.5 \text{ cm}^{-1}$) erhält man damit $\alpha(973.5 \text{ cm}^{-1}) \simeq 2\% \text{ cm}^{-1}$. Nimmt man an, daß bei Puls (b) nur ein Drittel der longitudinalen Moden anschwingt (diese Annahme wird in der noch folgenden Diskussion der Pulsform begründet!), folgt $\alpha(968.3 \text{ cm}^{-1}) \simeq 1.4\% \text{ cm}^{-1}$. Diese Werte sind in guter Übereinstimmung mit der von Taylor et al. /47/ berechneten Kleinsignalverstärkung (siehe Abschnitt I.1.1.3 und Abb. 3).

Zum Schluß sei noch kurz die Pulsform diskutiert: Im Gegensatz zu Puls (a), der eine nahezu glatte Form besitzt, ist die Struktur von Puls (b) durch zeitlich äquidistante Spitzen geprägt. Der zeitliche Abstand von etwa 15 nsec dieser Spitzen entspricht dabei genau der Umlaufzeit t_{um} der Photonen im Resonator. (Die Modulationstiefe von Puls (b) ist in Wirklichkeit stärker als in Abb. 19 gezeigt, da die endliche Bandbreite (125 MHz!) des zur Messung der Pulsstruktur verwendeten Transientenrekorders zu einer Verschleifung der Spitzenstruktur führt. Bei der Verwendung eines sehr schnellen Oszillographen (Tektronix 9407, Bandbreite 600 MHz) wurden auch Pulse mit nahezu 100% Modulationstiefe beobachtet!). Pulsstrukturen wie die von Puls (b) sind von TEA-CO₂-Lasern her bekannt /70/. Ursache ist eine Phasenkopplung zwischen longitudinalen Lasermoden infolge der nichtlinearen Verstärkung des aktiven Mediums bei hohen Strahlungsintensitäten ('self-mode-locking') /71/. Im allgemeinen wird dabei jedoch keine strenge Phasenkopplung zwischen allen innerhalb der Bandbreite der emittierten Strahlung enthaltenen Moden aufgebaut; vielmehr koppeln immer nur einige wenige benachbarte Moden /72/. Jede dieser Gruppen gekoppelter Moden erzeugt ihre eigene Pulsstruktur, die jedoch statistisch gegeneinander verschoben sind, so daß bei einer Vielzahl von Gruppen eine mehr oder weniger glatte Pulsstruktur entsteht. Dies ermöglicht den Schluß, daß bei einem Laserbetrieb weit oberhalb der Schwelle (Puls (a)) nahezu alle longitudinalen Moden innerhalb der Bandbreite der Laserstrahlung angeregt sind, während bei einem Betrieb nahe an der Schwelle (Puls (b)) wesentlich weniger Moden anschwingen. Es sei noch erwähnt, daß infolge der statistischen Ausbildung der Modenkopplung starke Unterschiede in der Pulsform aufeinanderfolgender Laserpulse auftraten. (Die räumliche Verteilung der Kleinsignalverstärkung im Entladungsvolumen ändert sich leicht von Entladung zu Entladung; damit sind auch die Voraussetzungen für die Kopplung der longitudinalen Moden von Puls zu Puls verschieden.) So wurde beispielsweise auch bei hoher Pulsenergie u.U. eine Struktur mit äquidistanten Spitzen beobachtet.

1.4.3 Breitbandemission

Um eine breitbandige Strahlungsemission zu erhalten, wurde die Kombination Gitter und Teleskop in Abb. 15 durch einen konkaven Molybdänspiegel mit Krümmungsradius $R = 4 \text{ m}$ ersetzt. Bei dem im folgenden berichteten Experiment beinhaltete der Laser nur wieder

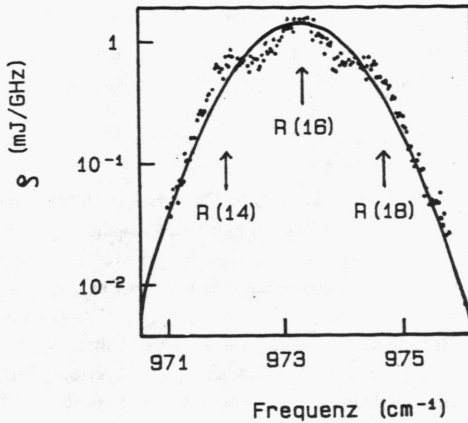


Abb. 20 Spektrale Energieverteilung der emittierten Laserstrahlung bei nichtfrequenzselektivem Resonator (Breitbandbetrieb).

eine Entladungsstrecke. Die gesamte Resonatorlänge betrug dann nur mehr 1.2 m. Die verwendete Gasmischung war $\text{CO}_2:\text{N}_2:\text{He} = 2:1:80$, der Gasdruck $p = 20$ atm.

Das Emissionsspektrum des Lasers unter dieser Konfiguration wurde mit Hilfe des in Abschnitt I.1.4.1 erwähnten Monochromators (Bandbreite $\Delta\nu_M = 1.5$ GHz) ermittelt. Es zeigte sich, daß die Emission nahezu vollständig im Zentrum des $10R$ -Zweiges erfolgte. Abb. 20 zeigt als Beispiel die spektrale Energieverteilung eines Laserpulses der Gesamtenergie von 150 mJ. Die Entladungsenergieichte war $15 \frac{\text{J}}{\text{atm}}$. Die geringfügige Modulation der Energieverteilung kennzeichnet die Zentren der druckverbreiterten Rotations-Schwingungs-Übergänge. Die Energieverteilung überdeckt somit den Bereich zwischen den Übergängen $R(14)$ und $R(18)$, d.h. etwa 3 cm^{-1} .

Die spektrale Energieverteilung wurde in Abb. 20 logarithmisch aufgetragen und erlaubt damit einen direkten Vergleich mit der von Taylor et al. /47/ berechneten Kleinsignalverstärkungskurve $\alpha(\nu)$ in Abb. 3. Unter Verwendung von Gl.(20), die die Zahl $Z(\nu)$ der Photonen in einer longitudinalen Mode der Frequenz ν nach der Aufbauzeit t_{auf} des Laserpulses angibt, gilt nämlich für die spektrale Energieverteilung:

$$\rho(\nu) = \frac{m Z(\nu) h \nu}{\Delta\nu_M} = \frac{m h \nu}{\Delta\nu_M} \exp \left((2l\alpha(\nu) - \beta) \frac{t_{\text{auf}}}{t_{\text{um}}} \right) \quad (22)$$

Hierbei ist m wieder die Zahl der longitudinalen Moden innerhalb der Bandbreite $\Delta\nu_M$ des Monochromators, der auf die Frequenz ν eingestellt ist, l die Länge des aktiven Volumens und β der Verlust pro Umlauf eines Photons. Nach Taylor et al. läßt sich die Kleinsignalverstärkung $\alpha(\nu)$ innerhalb des 10R-Zweiges annähernd durch eine Gaußsche Verteilung mit Zentrum bei 973.3 cm^{-1} (10R(16)) und der Halbwertsbreite von 21 cm^{-1} beschreiben (Abb.3). Weiterhin gilt nach Taylor et al. bis zu Entladungsenergiedichten von $40 \frac{\text{J}}{\text{l}_{\text{atm}}}$ ein linearer Zusammenhang zwischen Kleinsignalverstärkung im Zentrum des 10R-Zweiges und der Entladungsenergiedichte. Aus Abb.3 folgt somit für die maximale Kleinsignalverstärkung im 10R-Zweig für die hier vorliegende Entladungsenergiedichte von $15 \frac{\text{J}}{\text{l}_{\text{atm}}}$ ein Wert von 0.95% . Die anhand dieser Werte über Gl.(22) berechnete spektrale Energieverteilung ist in Abb.20 miteingezeichnet (durchgezogene Linie). Es zeigt sich eine sehr gute Übereinstimmung. Die restlichen Parameter in Gl.(22) waren dabei: $t_{\text{um}} = 8 \text{ nsec}$ (Resonatorlänge 1.2 m), $t_{\text{auf}} = 650 \text{ nsec}$ (gewonnen aus Messungen der Pulsstruktur wie in Abschnitt I.1.4.2), $\beta = 0.3$, $\Delta\nu_M = 1.5 \text{ GHz}$ und $m = 12$ (Frequenzabstand der longitudinalen Moden 125 MHz). Aus Gl. (22) erhält man als Halbwertsbreite (FWHM) der spektralen Energieverteilung der im breitbandigen Laserbetrieb emittierten Strahlung, die wegen der logarithmischen Darstellung aus Abb.20 schwer ablesbar ist, einen Wert von 2 cm^{-1} bzw. 60 GHz.

Von Untersuchungen des Breitbandbetriebs bei Verwendung von 2 Entladungsstrecken wurde abgesehen, da schon bei niederen Entladungsenergiedichten von $12.5 \frac{\text{J}}{\text{l}_{\text{atm}}}$, die einer Versorgungsspannung von 45 kV für den Marxgenerator entsprachen, Ausgangspulsenergien von über 300 mJ auftraten, die eine sofortige Zerstörung der NaCl-Brewsterfenster durch Plasmadurchbrüche an ihrer Oberfläche zur Folge hatten. Geringere Entladungsenergiedichten durch Reduktion der Versorgungsspannung des Marxgenerators einzustellen, war jedoch nicht möglich, da in diesem Fall die $\frac{E}{V}$ -Werte nicht mehr groß genug waren, um eine selbständige Gasentladung im gesamten Entladungsvolumen aufrechtzuerhalten. Es kam dann zu Bogenentladungen.

1.5 Zusammenstellung der Eigenschaften des Hochdruck-CO₂-Lasers und Ausbaumöglichkeiten

In Tab.6 sind einige wesentliche Eigenschaften des in dieser Arbeit entwickelten Hochdruck-CO₂-Lasers zusammengefaßt und mit den entsprechenden Werten anderer veröffentlichter Laserkonstruktionen verglichen. Hierbei wurden drei repräsentative Beispiele UV-vorionisierter Hochdruck-CO₂-Laser mit selbständiger Gasentladung ausgewählt. Laser mit anderen Anregungsmechanismen seien wegen des damit verbundenen technischen Aufwands nicht betrachtet. Der Vergleich zeigt, daß der hier vorgestellte Laser, insbesondere wenn man

die einfache und kompakte Konstruktion berücksichtigt, diverse Vorteile gegenüber den bisher veröffentlichten Hochdruck-CO₂-Lasern aufweist.

Als Ausblick seien noch drei Möglichkeiten angesprochen, die Effizienz des Lasers zu erhöhen:

- a) Wie bereits erwähnt, durfte die Ausgangspulsenergie 300 mJ nicht überschreiten, da andernfalls die Energiedichte der Strahlung an der Oberfläche der NaCl-Brewsterfenster so groß wurde, daß diese durch Plasmadurchbrüche beschädigt wurden. Ersetzt man jedoch das Strahlaufweitungsteleskop aus Molybdänspiegeln in Abb. 15 durch zwei analoge Teleskope, die innerhalb der Hochdruckkammer jeweils zwischen Fenster und Entladungsstrecke montiert sind und die ein Aufweitungsverhältnis 1:2 besitzen, lassen sich Pulsenergien >1 J erzielen, ohne daß es zu einer Schädigung der Fenster kommt. Problematisch ist dabei jedoch, daß es nach dem Füllen der Hochdruckkammer keine Möglichkeit mehr gibt, die Justierung der einzelnen Komponenten der Teleskope zu ändern; denn wie in Abschnitt I.1.3.5 diskutiert, beeinflusst insbesondere der Abstand der Spiegel eines Teleskops drastisch die Ausbildung der transversalen Resonatormoden.
- b) Eine Erhöhung des Elektrodenabstandes auf 1.5 cm würde die Wirkung der metallischen Elektroden als Wellenleiter für die Laserstrahlung reduzieren und die Ausbildung eines reinen TEM₀₀-Mode möglich machen. Dies erfordert jedoch neue Konstruktionen für die Komponenten des Entladungskreises (z.B. 3-stufiger Marxgenerator, bessere Hochspannungsdurchführungen in die Druckkammer etc.), da die Hochspannungspulse entsprechend dem Elektrodenabstand erhöht werden müßten.
- c) Wünschenswert wäre es auch, die spektrale Breite der emittierten Strahlung soweit zu reduzieren, daß nur mehr eine longitudinale Mode darin enthalten ist (SLM-Betrieb). Dies würde zeitlich vollständig glatte Pulsstrukturen und auch eine wesentliche Verbesserung der charakteristischen Eigenschaften des CH₃F-Raman-Lasers (siehe Abschnitt I.2.3) liefern. Eine Verringerung der Bandbreite erfordert zusätzliche dispersive Elemente im Resonator. Ein Beispiel wurde von Deka et al. /29/ demonstriert. Sie erreichten für einen 10 atm-CO₂-Laser eine Bandbreite von 400 MHz mit Hilfe eines aus drei Spiegeln bestehenden Resonators (ein dritter zusätzlicher teilreflektierender Spiegel zwischen Gitter und Hochdruckkammer wirkt zusammen mit dem Gitter als ein Fabry-Perot-Interferometer). Eine andere Möglichkeit wäre der Einbau von Fabry-Perot-Etalons im Resonator. Bernard et al. /73/ erzielten mit einem Ge-Etalon bei einem TEA-CO₂-Laser eine Verringerung der Bandbreite, die üblicherweise auch 2...4 GHz beträgt, auf 200 MHz. Erste Versuche mit einem Ge-Etalon, im vorliegenden Laser eine Reduktion der Bandbreite durchzuführen, scheiterten an der relativ hohen Absorption der Laserstrahlung in dem 1.3 cm dicken Ge-Etalon - der Absorptionskoeffizient von Ge für 10.6 μm -Strahlung beträgt $\alpha = 0.045 \text{ cm}^{-1}$ /74/. Der Einsatz von ZnSe-Etalons ($\alpha = 0.005 \text{ cm}^{-1}$ /74/) wäre daher empfehlenswert. Um einen echten SLM-Betrieb zu erhalten, müßte u.U. eine Kombination mehrerer Etalons eingebaut werden.

	Mathieu et al. /16,57/	Danly et al. /17,63/	Carman et al. /27/	Werling et al. /61,62/
Gasmischung (CO ₂ :N ₂ :He)	1:1:8	1:1:8	1:1:8	3:1:60
Druck (Durchfluß)	10-12 atm (1-2 l/min)	10-12 atm (10-15 l/min)	8-12 atm (2 l/min)	20-25 atm (sealed off)
Vorionisierungs- anordnung	seriell geschaltete Funkenstrecken eigene Hochspan- nungsversorgung 40 kV/8 J wird 0.4 µsec vor der Hauptentlad- ung gezündet	seriell geschaltete Funkenstrecken eigene Hochspan- nungsversorgung 50 kV/12 J wird 0.5 µsec vor der Hauptentlad- ung gezündet	seriell geschaltete Funkenstrecken eigene Hochspan- nungsversorgung 25 kV/16 J wird 2 µsec vor der Hauptentlad- ung gezündet	parallel geschal- tete Funkenstrecken über 70 pF - Kon- densatoren direkt an die Hauptent- ladung gekoppelt
Elektrodenform	Chang	Rogowski		Chang
aktives Volumen (H×B×L)	0.8×0.8×27 cm ³	0.7×0.64×35 cm ³	0.6×0.9×16 cm ³	1×1×80 cm ³
Spannungsver- sorgung und typ. Ladespannung	2-stufige Marx- bank 3 nF pro Stufe 70 kV	5-stufige Marx- bank 16.2 nF pro Stufe 20 kV	LC-Verdoppler- schaltung 4.8 nF pro Stufe 45 kV	2-stufige Marx- bank 10 nF pro Stufe 60 kV
typ. E/N-Wert	7 · 10 ⁻¹⁶ Vcm ²	6 · 10 ⁻¹⁶ Vcm ²	6 · 10 ⁻¹⁶ Vcm ²	2.5 · 10 ⁻¹⁶ Vcm ²
typ. Entladungs- energiedichte	100 $\frac{\text{J}}{\text{l atm}}$	100 $\frac{\text{J}}{\text{l atm}}$	140 $\frac{\text{J}}{\text{l atm}}$	22 $\frac{\text{J}}{\text{l atm}}$
Abstimm- bereich	9R(10)- 9R(32) 9P(10)- 9P(32) 10R(8)-10R(32) 10P(8)-10P(32)	abstimmbar im 9R-/10R-Zweig nicht in den P-Zweigen	10R(10)-10R(28) 10P(10)-10P(26)	9R(6)- 9R(28) 9P(8)- 9P(28) 10R(6)-10R(30) 10P(10)-10P(26)
Ausgangspuls- energie bei 10R(16)	150 mJ	80-120 mJ	60-80 mJ	300 mJ
Bandbreite (FWHM)	3.6 GHz	4 GHz	4.2 GHz	4.8 GHz
Pulsdauer	50 nsec	100 nsec	50 nsec	130 nsec
Wiederholrate	0.5 Hz	0.3-2 Hz		0.5 Hz
Auskoppel- spiegel	Ge 50% Refl.	Ge 65% Refl.	Ge 80% Refl. eben	Ge 70% Refl. R = 25 m
Gitter	150 l/mm	150 l/mm	150 l/mm	150 l/mm
Teleskop	1:3	1:2.5		1:3

Tab. 6 Vergleich der wesentlichen Eigenschaften bisher veröffentlichter Hochdruck-CO₂-Laser mit denen des in dieser Arbeit entwickelten Lasers.

I.2. Der CH_3F -Raman-Laser

Die Erzeugung kohärenter FIR-Strahlung durch optisches Pumpen von molekularen Gasen kann entweder durch stimulierte resonante Übergänge zwischen Molekülniveaus oder durch den stimulierten Ramaneffekt erfolgen. Die beiden Möglichkeiten sind in Abb. 21 skizziert. Im ersten Fall erfolgt die Erzeugung der FIR-Strahlung in zwei Schritten: durch resonante Absorption der Pumpstrahlung wird eine Überbesetzung eines Rotationszustandes in einem angeregten Schwingungszustand erzeugt. Kohärente FIR-Strahlung entsteht dann durch stimulierte Übergänge zwischen dem überbesetzten Rotationsniveau und seinem Nachbarniveau. Die Frequenz der FIR-Strahlung ist dabei durch den Energieabstand der Rotationsniveaus fest vorgegeben. Beim Ramaneffekt handelt es sich um einen Zweiphotonenprozeß: das Molekül absorbiert ein Pumpphoton und emittiert simultan ein FIR-Photon. Die Differenz der Energie der beiden Photonen wird zur Anregung des Moleküls verwendet. Bei hohen Intensitäten der Pumpstrahlung verläuft dieser Prozeß stimuliert. In diesem Fall ist keine Übereinstimmung der beiden Frequenzen mit Frequenzen resonanter Übergänge im Molekül erforderlich; nur die Differenzfrequenz entspricht einer resonanten Anregung des Moleküls. Der stimulierte Ramaneffekt erlaubt damit die abstimmbare Erzeugung kohärenter FIR-Strahlung, wenn eine abstimmbare Pumpquelle zur Verfügung steht.

Methylfluorid (CH_3F) bietet beide Möglichkeiten für die Erzeugung kohärenter FIR-Strahlung. So war der erste optisch gepumpte FIR-Laser (Chang et al. /7/) ein von einem CO_2 -Laser gepumpter CH_3F -Laser des ersten Typs: die resonante Absorption der $9P(20)$ -Strahlung des CO_2 -Lasers führt zur Emission der bekannten $496.1 \mu\text{m}$ Strahlung von CH_3F . FIR-Erzeugung durch stimulierte Ramanemission in CH_3F wurde von Biron et al. /15/ gezeigt. Obwohl sie infolge des verwendeten TEA- CO_2 -Lasers die FIR-Strahlung nicht abstimmbare erzeugen konnten, war ihre Arbeit ein wichtiger Schritt für die Entwicklung abstimmbarer FIR-Laser. Sie demonstrierten erstmals, daß bei genügend starker Pumpintensität eine stimulierte Ramanstreuung in CH_3F möglich ist, auch wenn die Pumpstrahlungsfrequenz bis zu 30 GHz von der Frequenz eines Rotations-Schwingungs-Übergangs in CH_3F abweicht. Dies ist die größte bei FIR-Lasern bisher beobachtete Abweichung, die üblicherweise bei 2 GHz liegt /63/. Die Ergebnisse von Biron et al. wurden in den letzten Jahren von mehreren Gruppen zur Erzeugung kontinuierlich abstimmbarer FIR-Strahlung ausgeweitet, indem ein Hochdruck- CO_2 -Laser als Pumpquelle verwendet wurde. Danly et al. /17/ demonstrierten eine abstimmbare Erzeugung von FIR-Strahlung im Bereich von $33 \text{ cm}^{-1} - 40 \text{ cm}^{-1}$, Mathieu et al. /16/ sogar im Bereich von $25 \text{ cm}^{-1} - 45 \text{ cm}^{-1}$. Izatt et al. zeigten, daß FIR-Strahlung auch bei kürzeren Wellenlängen, nämlich in einer Reihe von Frequenzintervallen zwischen 57 cm^{-1} und 70 cm^{-1} , durch stimulierte Ramanstreuung

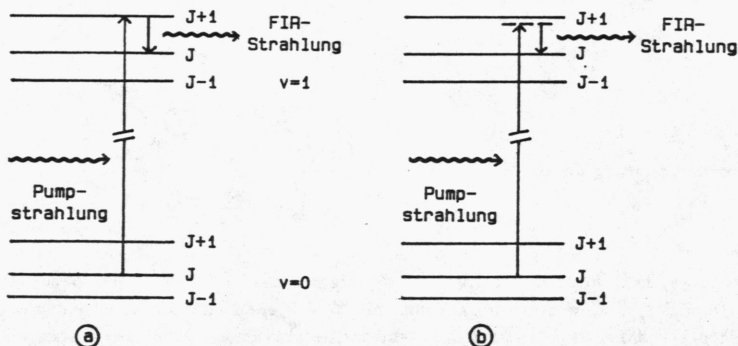


Abb. 21 Erzeugung von FIR-Strahlung durch einen resonanten Zweistufenprozess (a) und stimulierten Ramanereffekt (b).

in dem isotopischen Molekül $^{13}\text{CH}_3\text{F}$ kontinuierlich erzeugt werden kann /75/. (Die Eigenschaften der von ihnen verwendeten Hochdruck- CO_2 -Laser sind in Tab. 6 zusammengefaßt.)

In diesem Kapitel wird erstmals die vollständige Emissionscharakteristik eines von einem 20 atm- CO_2 -Laser gepumpten CH_3F -Raman-Lasers vorgestellt, die von 23 cm^{-1} bis 45 cm^{-1} reicht, und die Effizienz des Lasers untersucht /76,77/. Bevor nun der Aufbau des Lasers und seine charakteristischen Eigenschaften geschildert werden, seien kurz die Grundlagen der abstimmbaren Erzeugung von FIR-Strahlung durch stimulierte Ramanstreuung in CH_3F besprochen.

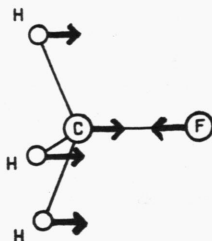
2.1 Stimulierte Ramanemission in CH_3F

2.1.1 Energieniveauschema eines CH_3F -Moleküls

CH_3F ist ein symmetrisches Kreismolekül mit C_{3v} -Symmetrie und besitzt 6 verschiedene Normalschwingungen, die alle sowohl infrarotaktiv als auch ramanaktiv sind /36/. Für die Diskussion in diesem Abschnitt ist von ihnen nur die ν_3 -Normalschwingung, die sog. C-F-Streckschwingung, die in Abb. 22 skizziert ist, relevant. Ihre Grundfrequenz beträgt $\nu_3 = 1048.62\text{ cm}^{-1}$. Das oszillierende Dipolmoment zeigt in Richtung der Figurenachse.

Zur Beschreibung der Rotationszustände eines CH_3F -Moleküls sind 3 Quantenzahlen nötig, die üblicherweise mit J , K und M bezeichnet werden. J beschreibt den Betrag des Drehim-

Abb. 22

 ν_3 -Normalschwingung von CH_3F .

pulses, K seine Komponente in Richtung der Figurenachse. Die möglichen Werte von K sind $K = 0, 1, \dots, J$, wobei jeder Zustand mit $K \neq 0$ zweifach entartet ist entsprechend der Möglichkeit der Rotation um die Figurenachse in oder entgegen dem Uhrzeigersinn. M ($-2J \leq M \leq 2J$) bezeichnet letztendlich die verschiedenen Einstellmöglichkeiten des Gesamtdrehimpulses bezüglich einer raumfesten Achse. Solange eine solche nicht ausgezeichnet ist, ist jeder Zustand $|J, K\rangle$ $2(2J + 1)$ -fach entartet, falls $K \neq 0$ bzw. $(2J + 1)$ -fach entartet, falls $K = 0$. Die Gesamtenergie eines Rotationszustandes $|J, K\rangle$ errechnet sich nach

$$E_{JK}^v = h[B^v J(J+1) + (A^v - B^v)K^2 - D_J^v J^2(J+1)^2 - D_{JK}^v J(J+1)K^2 - D_K^v K^4] \quad (23)$$

A und B sind dabei die Rotationskonstanten; sie hängen vom Trägheitsmoment für die Rotation um die Symmetrieachse bzw. um eine Achse senkrecht dazu ab (es gilt $A - B > 0$, d.h. CH_3F ist ein prolates Molekül), die Konstanten D_J, D_{JK} und D_K berücksichtigen die Zentrifugalaufweitung. Die Kennzeichnung der Konstanten mit dem Index v berücksichtigt, daß sie für die verschiedenen Zustände $|v\rangle$ der ν_3 -Schwingung unterschiedliche Größen besitzen. Tab. 7 gibt die Werte der Konstanten für die Moleküle $^{12}\text{CH}_3\text{F}$ und $^{13}\text{CH}_3\text{F}$ an. Die Gesamtenergie eines Zustands $|v, J, K\rangle$ setzt sich aus Schwingungs- und Rotationsenergie zusammen: $E_{vJK} = h(v + \frac{1}{2})\nu_3 + E_{JK}^v$.

	$^{12}\text{CH}_3\text{F}$	$^{13}\text{CH}_3\text{F}$
ν_3	31436.558 GHz	30803.473 GHz
B^0	25536.15 MHz	24862.79 MHz
B^1	25197.56 MHz	24542.27 MHz
$(A^1 - B^1) - (A^0 - B^0)$	44.1 MHz	31.36 MHz
D_J^0	60.26 KHz	59.42 KHz
D_J^1	57.43 KHz	56.27 KHz
D_{JK}^0	439.46 KHz	425.18 KHz
D_{JK}^1	518.22 KHz	478.91 KHz
$(D_K^1 - D_K^0)$	-101.35 KHz	-69.46 KHz

Tab. 7 Molekülkonstanten für CH_3F nach /78/.

Die relative Besetzung $\rho_{JK}^v = \frac{N_{JK}^v}{N^v}$ des Rotationszustandes $|J, K\rangle$ bezüglich der Besetzung N^v des zugehörigen Schwingungszustandes $|v\rangle$ errechnet sich nach der Boltzmannverteilung unter Berücksichtigung des statistischen Gewichts g_{JK} des Zustands $|J, K\rangle$:

$$\rho_{JK}^v \sim g_{JK} \exp\left(-\frac{E_{JK}^v}{kT}\right) \quad (24a)$$

Hierbei gilt:

$$g_{JK} = \begin{cases} 8(2J+1) & K \bmod 3 = 0; K \neq 0 \\ 4(2J+1) & \text{sonst} \end{cases} \quad (24b)$$

Dies folgt aus dem oben erwähnten Entartungsgrad des Niveaus $|J, K\rangle$ zusammen mit Symmetriebetrachtungen über den Austausch der drei identischen Wasserstoffkerne (Kernspin $I = \frac{1}{2}$) innerhalb des CH_3F -Moleküls [36]. Die relativen Besetzungen $\rho_{JK}^{v=0}$ sind für verschiedene festgehaltene K -Werte als Funktion von J in Abb. 23 aufgetragen für eine Temperatur $T = 300$ K. Man sieht, daß die Zustände mit $K = 3$ die größte relative Besetzung besitzen; sie werden daher den größten Beitrag zur Kleinsignalverstärkung für die stimulierte Ramanstreuung liefern. Das Maximum der einzelnen Besetzungsverteilungen liegt etwa bei $J = 11$.

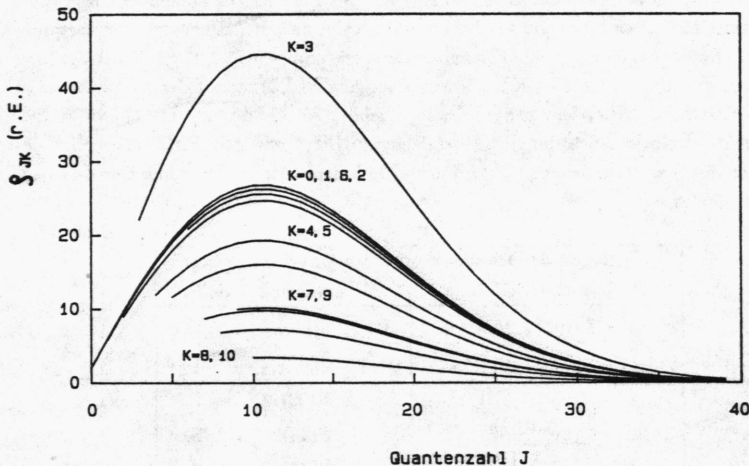


Abb. 23 Relative Besetzung der Rotationszustände des Schwingungsgrundzustandes in CH_3F .

Die Auswahlregeln für die Änderung der Rotationsquantenzahlen bei Rotations-Schwingungs-Übergängen lauten:

$$\Delta J = 0, \pm 1 \quad \Delta K = 0 \quad \text{für } K \neq 0$$

$$\Delta J = \pm 1 \quad \Delta K = 0 \quad \text{für } K = 0$$

Die Übergänge mit $\Delta J = 1$ bilden den sog. *R*-Zweig, die Übergänge mit $\Delta J = -1$ den *P*-Zweig (analog zum CO_2 -Molekül). Der zusätzliche Zweig mit $\Delta J = 0$ heißt *Q*-Zweig. Ein Übergang wird dabei wieder charakterisiert durch die Quantenzahlen des tieferliegenden Niveaus. Aus Gl.(23) errechnet man beispielsweise die Frequenz des Übergangs $|0, J, K\rangle \rightarrow |1, J+1, K\rangle$:

$$\begin{aligned} R(J, K) = \nu_3 + K^2[(A^1 - B^1) - (A^0 - B^0)] - K^4(D_K^1 - D_K^0) \\ + (J+1)[(B^1 + B^0) - K^2(D_{JK}^1 + D_{JK}^0)] \\ + (J+1)^2[(B^1 - B^0) - K^2(D_{JK}^1 - D_{JK}^0) - (D_J - D_J^0)] \\ - 2(J+1)^3[D_J^0 + D_J^1] - (J+1)^4[D_J - D_J^0] \end{aligned} \quad (25)$$

Einsetzen der Konstanten aus Tab. 7 zeigt, daß man für ein gegebenes *J* eine Vielzahl eng benachbarter Übergänge entsprechend den zu *J* gehörenden Quantenzahlen *K* erhält. Ihre Frequenzen sind alle innerhalb eines Intervalls von $\pm 0.01 \text{ cm}^{-1}$ um die Frequenz des Übergangs $R(J, 3)$ verteilt, der ja nach Abb. 23 den dominanten Beitrag aller Übergänge $R(J, K)$ liefert. In Abschnitt I.2.3.1 wird gezeigt, daß bei dem hier vorgestellten Lasersystem zur abstimmbaren FIR-Erzeugung der minimal erkennbare Frequenzunterschied 0.04 cm^{-1} beträgt. Man kann daher alle Übergänge $R(J, K)$ mit dem Übergang $R(J, 3)$ identifizieren. Dieser sei in Zukunft einfach mit $R(J)$ bezeichnet. Entsprechendes gilt für den *P*- und *Q*-Zweig. Der Überlapp von Rotations-Schwingungs-Übergängen in CH_3F und dem Emissionsbereich des im vorigen Kapitel beschriebenen Hochdruck- CO_2 -Lasers (Abb. 16b) ist in Tab. 8 zusammengefaßt.

Emissionsbereich des Hochdruck- CO_2 -Lasers	Rotations-Schwingungs- Übergänge in CH_3F
9R: $1067 \text{ cm}^{-1} - 1087 \text{ cm}^{-1}$	R(13) - R(26) $^{12}\text{CH}_3\text{F}$
9P: $1037 \text{ cm}^{-1} - 1059 \text{ cm}^{-1}$	Q(1) - Q(30) $^{12}\text{CH}_3\text{F}$ R(6) - R(20) $^{13}\text{CH}_3\text{F}$
10R: $964 \text{ cm}^{-1} - 984 \text{ cm}^{-1}$	P(32) - P(39) $^{12}\text{CH}_3\text{F}$ P(23) - P(32) $^{13}\text{CH}_3\text{F}$
10P: $932 \text{ cm}^{-1} - 956 \text{ cm}^{-1}$	—

Tab. 8 Rotations-Schwingungs-Übergänge in CH_3F innerhalb des Emissionsbereichs des Hochdruck- CO_2 -Lasers.

Zum Schluß seien noch die wichtigsten Relaxationsmechanismen in CH_3F besprochen. Wie schon beim CO_2 -Molekül diskutiert, ist auch beim CH_3F -Molekül der energetische Abstand benachbarter Rotationsniveaus klein gegen die kinetische Translationsenergie der Moleküle bei einer Temperatur $T = 300$ K. Durch Stöße der Moleküle untereinander erfolgt daher ein schneller Austausch von Rotations- und Translationsenergie der Moleküle (R-T-Transfer). Die charakteristische Relaxationszeit τ_{RT} beträgt nach Trappeniers /79/: $\tau_{RT} \cdot p = 8 \text{ nsec} \cdot \text{torr}$ für kleine J ($J \leq 20$). Mit zunehmendem J nimmt τ_{RT} jedoch zu. Das Energiegleichgewicht innerhalb des ν_3 -Schwingungssystems stellt sich durch den Austausch von Schwingungsenergie infolge Stöße der Moleküle untereinander (V-V-Transfer) in einer Zeit τ_{VV} ein, für die gilt: $\tau_{VV} \cdot p = 5 \mu\text{sec} \cdot \text{torr} / 63/$. Der Austausch von Vibrationsenergie und Translationsenergie der Moleküle durch Stöße (V-T-Transfer) erfolgt wesentlich langsamer. Die Zeitkonstante τ_{VT} für diesen Prozeß beträgt $\tau_{VT} \cdot p = 1.7 \text{ msec} \cdot \text{torr} / 80/$. Bei einem Betriebsdruck von $p = 17$ Torr erhält man somit die Zeiten:

$$\tau_{RT} = 0.5 \text{ nsec} \quad \tau_{VV} = 300 \text{ nsec} \quad \tau_{VT} = 100 \mu\text{sec}$$

Da die Dopplerverbreiterung in CH_3F bei $T = 300$ K nur einige MHz beträgt /39/, ist die Stoßverbreiterung der Prozeß, der die Linienbreite eines Rotations-Schwingungs-Übergangs bestimmt. Die Linienform der Übergänge ist damit durch eine Lorentzform beschreibbar. Für das obige Beispiel errechnet man aus τ_{RT} mit Hilfe von Gl.(4) eine Halbwertsbreite (FWHM) von $\Delta\nu = 0.02 \text{ cm}^{-1}$ für die Linien $R(J)$ mit $J \leq 20$. Für $J > 20$ nimmt die Linienbreite ab.

2.1.2 Kleinsignalverstärkung für die stimulierte Ramanstreuung in CH_3F

Bei der Ramanstreuung eines Photons (Frequenz ν_p , Wellenvektor \vec{k}_p) an einem Molekül, das einen ramanaktiven Übergang zwischen 2 Niveaus $|1\rangle, |2\rangle$ (ihre Energiedifferenz sei $h\nu_{21}$) besitzt, wird dieses Photon vernichtet, während gleichzeitig ein gestreutes Photon (Frequenz ν_s , Wellenvektor \vec{k}_s) entsteht und das Molekül in den angeregten Zustand $|2\rangle$ übergeht. Hierbei gilt $\nu_s = \nu_p - \nu_{21}$. Läßt man Laserstrahlung der Frequenz ν_p und Intensität I_p in z -Richtung auf eine Zelle mit Volumen V , die N ramanaktive Moleküle enthalten möge, fallen, gilt für die Intensität I_s des gestreuten Lichts der Frequenz ν_s , wenn man von allen möglichen Streuprozessen nur die mit $\vec{k}_s \parallel \vec{k}_p \parallel \hat{z}$ betrachtet:

$$\frac{dI_s}{dz} = g(\nu_p) I_p \left(I_s + \frac{h\nu_s c}{V} \right) \quad /81, 82/ \quad (26)$$

c ist die Lichtgeschwindigkeit (der Brechungsindex von CH_3F kann $n = 1$ gesetzt werden!), $g(\nu_p)$ ist ein Maß für die Stärke der inelastischen Streuung von Photonen der Frequenz ν_p

am Übergang $|1\rangle \rightarrow |2\rangle$. Mit Hilfe der Störungstheorie 2.Ordnung erhält man für $g(\nu_p)$ /81,82/:

$$g(\nu_p) = \frac{1}{c^2 \varepsilon_0^2 \hbar^3} \frac{\nu_s}{\Delta\nu} \frac{N}{V} < \left| \frac{1}{2\pi} \sum_j \left\{ \frac{(\vec{e}_s \vec{\mu}_{2j})(\vec{e}_p \vec{\mu}_{j1})}{\nu_p - \nu_{j1}} + \frac{(\vec{e}_p \vec{\mu}_{2j})(\vec{e}_s \vec{\mu}_{j1})}{\nu_s + \nu_{j1}} \right\} \right|^2 > \quad (27)$$

Hierbei ist $\Delta\nu$ die Breite (FWHM) des Übergangs $|1\rangle \rightarrow |2\rangle$. Die Summation erfolgt über alle möglichen Eigenzustände des Moleküls. \vec{e}_s und \vec{e}_p seien die Polarisationsrichtungen der einfallenden und der gestreuten Strahlung, $\vec{\mu}_{ab}$ bezeichne das Dipolmatrixelement für den Übergang $|a\rangle \rightarrow |b\rangle$ ($\nu_{ab} = \frac{1}{\hbar}(E_b - E_a)$). Die Klammern $< \dots >$ bedeuten eine Mittelung über die statistisch verteilten Orientierungen der Moleküle im Gas, die sich in den Dipolmatrixelementen widerspiegeln.

Im Falle kleiner Pumpintensitäten I_p ist die Intensität der in die Mode ν_s, \vec{k}_s gestreuten Strahlung sehr gering ($I_s \ll \frac{\hbar\nu_p}{V}$). Gl.(26) beschreibt dann die spontane Ramanstreuung. Bei sehr starken Pumpintensitäten I_p wird jedoch auch die Intensität der gestreuten Strahlung I_s groß. Gilt $I_s \gg \frac{\hbar\nu_p}{V}$, hat Gl.(26) die exponentielle Lösung

$$I_s(z) = I_s(z=0) e^{g(\nu_p) I_p z} \quad (28)$$

Man spricht in diesem Fall von stimulierter Ramanstreuung. Die Größe $\alpha(\nu_p) = g(\nu_p) I_p$ heißt die zugehörige Kleinsignalverstärkung. Sie ist abhängig von der Pumpintensität I_p .

Abb.24 a zeigt die schematische Darstellung für einen Ramanprozeß an einem Schwingungsübergang $|1\rangle \rightarrow |2\rangle$ in einem Molekül. Liegt dabei die Pumpfrequenz ν_p in der Nähe eines resonanten Übergangs, der zudem noch infrarotaktiv ist (Übergang $|1\rangle \rightarrow |3\rangle$), können in Gl.(27) alle Summanden bis auf $\frac{(\vec{e}_s \vec{\mu}_{23})(\vec{e}_p \vec{\mu}_{31})}{\nu_p - \nu_{31}}$ ignoriert werden, da dieser sowohl infolge des sehr klein werdenden Nenners als auch infolge der nicht verschwindenden Dipolmatrixelemente des Zählers (auch der Übergang $|3\rangle \rightarrow |2\rangle$ sei als infrarotaktiv angenommen!) den dominanten Beitrag in der Summe liefert. Der Ramanprozeß wird in diesem Fall als quasiresonant bezeichnet. Für die Kleinsignalverstärkung erhält man aus (27):

$$\alpha(\nu_p) = \frac{1}{c^2 \varepsilon_0^2 \hbar^3} \frac{\nu_s}{\Delta\nu} \frac{N}{V} < \left| \frac{1}{2\pi} \frac{(\vec{e}_s \vec{\mu}_{23})(\vec{e}_p \vec{\mu}_{31})}{\nu_p - \nu_{31}} \right|^2 > I_p \quad (29)$$

Sie ist somit symmetrisch um die Frequenz ν_{31} des infrarotaktiven Übergangs $|1\rangle \rightarrow |3\rangle$ und nimmt $\sim \frac{1}{\delta p^2}$ ($\delta p = \nu_p - \nu_{31}$) ab. (Gl.(29) gilt natürlich nicht für $\nu_p = \nu_{31}$!).

Nahezu alle FIR-Laser, die auf der Basis der stimulierten Ramanstreuung arbeiten, können durch den quasiresonanten Ramanprozeß aus Abb. 24 a beschrieben werden. Denn bei den meisten molekularen Gasen wurde eine stimulierte Ramanemission nur beobachtet, wenn die Frequenz des Pumplasers (üblicherweise ein TEA-CO₂-Laser) in unmittelbarer Nähe

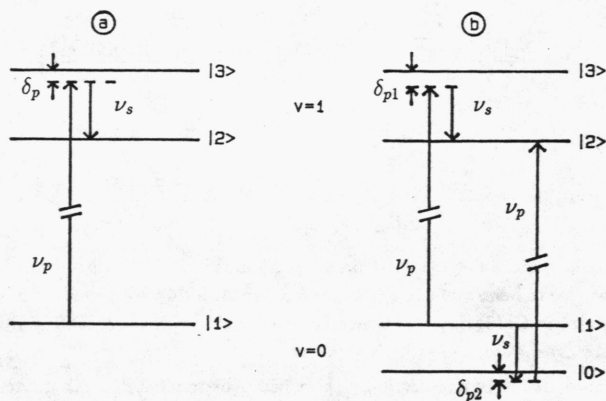


Abb. 24 Schematische Darstellung für einen einfachen und doppelten quasiresonanten Ramanprozeß an einem Schwingungsübergang in einem Molekül.
 $v = 0$ bezeichne den Schwingungsgrundzustand, $v = 1$ den ersten angeregten Zustand; die Niveaus $|0\rangle, |1\rangle$ bzw. $|2\rangle, |3\rangle$ seien ein Teil der zugehörigen Rotationszustände.

eines resonanten Übergangs (Abweichungen $\leq 2 \text{ GHz} / 63/$) lag. Die anderen Übergänge im Molekül waren dann soweit von der Pumpfrequenz entfernt, daß 3 Molekülniveaus zur Beschreibung der Lasertätigkeit ausreichten.

Bei größeren Abweichungen der Pumpfrequenz ν_p von der des resonanten Übergangs $|1\rangle \rightarrow |3\rangle$, wie sie von Biron et al. /15/ in CH_3F beobachtet wurden ($\delta_p \sim 30 \text{ GHz}$), gilt diese Aussage nicht mehr. Dann kann nämlich für einen zusätzlichen Rotationszustand $|0\rangle$ im Schwingungsgrundzustand der Übergang $|0\rangle \rightarrow |2\rangle$ infrarotaktiv sein und eine Frequenz, die ebenfalls nahe an der Pumpfrequenz ν_p liegt, besitzen (Abb. 24 b). Ein zweiter quasiresonanter Ramanprozeß wird dann wirksam, bei dem das gestreute Photon sozusagen im Schwingungsgrundzustand emittiert wird. Da bei beiden Prozessen das Molekül vom Zustand $|1\rangle$ in den Zustand $|2\rangle$ übergeht, ist die Frequenz des gestreuten Photons in beiden Fällen die gleiche, nämlich $\nu_s = \nu_p - \nu_{21}$. Beide Prozesse tragen damit zur Kleinsignalverstärkung für die stimulierte Streuung von Photonen der Frequenz ν_p in solche der Frequenz ν_s bei. Man spricht von einem doppelten quasiresonanten Ramanprozeß. Da die Wahrscheinlichkeitsamplituden für die beiden Prozesse bei der Berechnung der Kleinsignalverstärkung zu addieren

sind, erhält man

$$\alpha(\nu_p) = \frac{1}{c^2 \varepsilon_0^2 \hbar^3} \frac{\nu_s}{\Delta \nu} \frac{N}{V} < \left| \frac{1}{2\pi} \left\{ \frac{(\vec{e}_s \vec{\mu}_{23})(\vec{e}_p \vec{\mu}_{31})}{\nu_p - \nu_{31}} + \frac{(\vec{e}_p \vec{\mu}_{20})(\vec{e}_s \vec{\mu}_{01})}{\nu_{20} - \nu_p} \right\} \right|^2 > I_p \quad (30)$$

Mit $\delta p_1 = \nu_p - \nu_{31}$, $\delta p_2 = \nu_p - \nu_{20}$ und $\tau_{RT} = \frac{1}{\pi \Delta \nu}$ (R-T-Relaxationszeit) folgt

$$\alpha(\nu_p) = \frac{\pi \tau_{RT} \nu_s}{c^2 \varepsilon_0^2 \hbar^3} \frac{N}{V} < \left| \frac{1}{2\pi} \left\{ \frac{(\vec{e}_s \vec{\mu}_{23})(\vec{e}_p \vec{\mu}_{31})}{\delta p_1} - \frac{(\vec{e}_p \vec{\mu}_{20})(\vec{e}_s \vec{\mu}_{01})}{\delta p_2} \right\} \right|^2 > I_p \quad (31)$$

Die Kleinsignalverstärkung $\alpha(\nu_p)$ ist damit nicht mehr symmetrisch um die Frequenz ν_{31} bzw. ν_{20} , wie sie es beim einfachen Ramanprozeß in Abb. 24 a war, sondern zeigt durch den Interferenzterm in Gl.(31) ein wesentlich komplizierteres Verhalten. Dies soll im folgenden genauer betrachtet werden.

In CH_3F wird der Ausgangszustand $|1\rangle$ gemäß Abschnitt I.2.1.1 charakterisiert durch ein Tripel von Quantenzahlen $|1\rangle = |0, J, K\rangle$. Unter der Berücksichtigung, daß die infrarotaktiven Rotations-Schwingungs-Übergänge $|1\rangle \rightarrow |3\rangle, |0\rangle \rightarrow |2\rangle$ den Auswahlregeln $\Delta J = 0, \pm 1, \Delta K = 0$ und die infrarotaktiven reinen Rotationsübergänge $|3\rangle \rightarrow |2\rangle$ bzw. $|1\rangle \rightarrow |0\rangle$ der Auswahlregel $\Delta J = -1, \Delta K = 0$ gehorchen müssen, gibt es bei gegebenem Ausgangszustand $|1\rangle = |0, J, K\rangle$ 3 verschiedene Sätze für die Zustände $|0\rangle, |1\rangle, |2\rangle, |3\rangle$. Dies sind:

$$\begin{aligned} \Delta J = +1 & \quad |0\rangle = |0, J-1, K\rangle \quad |1\rangle = |0, J, K\rangle \quad |2\rangle = |1, J, K\rangle \quad |3\rangle = |1, J+1, K\rangle \\ \Delta J = -1 & \quad |0\rangle = |0, J-1, K\rangle \quad |1\rangle = |0, J, K\rangle \quad |2\rangle = |1, J-2, K\rangle \quad |3\rangle = |1, J-1, K\rangle \\ \Delta J = 0 & \quad |0\rangle = |0, J-1, K\rangle \quad |1\rangle = |0, J, K\rangle \quad |2\rangle = |1, J-1, K\rangle \quad |3\rangle = |1, J, K\rangle \end{aligned}$$

Je nachdem, ob die resonanten Rotations-Schwingungs-Übergänge, in deren Nähe die Pumpfrequenz ν_p liegt, der Auswahlregel $\Delta J = 1$, $\Delta J = -1$ oder $\Delta J = 0$ gehorchen, spricht man von *R-Zweig*-, *P-Zweig*- oder *Q-Zweig*-Pumpen. Die Kleinsignalverstärkung läßt sich nach Danly et al. /63/ in der folgenden Form zusammenfassen:

$$\alpha_{JK}(\nu_p) = \frac{\pi \tau_{RT} \nu_s}{c^2 \varepsilon_0^2 \hbar^3} \frac{N}{V} \rho_{JK}^0 \left(A_{JK} \frac{\mu_{IR}^2 \mu_o^2}{\delta p_1^2} + B_{JK} \frac{\mu_{IR}^2 \mu_o^2}{\delta p_2^2} - C_{JK} \frac{\mu_{IR}^2 \mu_o^2}{\delta p_1 \delta p_2} \right) \cdot I_p \quad (32)$$

Die Koeffizienten A_{JK}, B_{JK} und C_{JK} sind für jeden der erwähnten 3 Fälle verschieden. Sie beinhalten neben der Mittelung über die statistisch verteilten Orientierungen der Moleküle auch die relative Polarisation von Pumpstrahlung und gestreuter FIR-Strahlung. Nach Danly et al. /63/ erhält man maximale Verstärkung beim *R*- bzw. *P*-Zweig-Pumpen, wenn Pumpstrahlung und gestreute Strahlung parallel polarisiert sind, und beim *Q*-Zweig-Pumpen, wenn sie senkrecht zueinander polarisiert sind. Die Koeffizienten sind für diese

TABLE I
GAIN COEFFICIENTS FOR SYMMETRIC-TOP MOLECULES

	A_{JK}	B_{JK}	C_{JK}
$R(0)$	$\frac{(J+1)^2 - K^2(4J^2 + 8J + 5)}{15(J+1)^2(2J+1)^2(2J+3)}$	$\frac{(J^2 - K^2)(4J^2 + 1)}{15J^2(2J+1)^2(2J-1)}$	$\frac{4(J^2 - K^2)(J+1)^2 - K^2}{15J(J+1)(2J+1)^2}$
$Q(\perp)$	$\frac{(J^2 - K^2)K^2(4J+1)}{30J^2(J+1)(2J+1)}$	$\frac{(J^2 - K^2)K^2(4J-1)}{30J^2(2J+1)(2J-1)}$	$\frac{(J^2 - K^2)K^2(4J+1)}{15J^2(J-1)(2J+1)}$
$P(0)$	$\frac{2(J^2 - K^2)((J-1)^2 - K^2)}{15J(2J+1)(J-1)(2J-1)}$	$\frac{2(J^2 - K^2)((J-1)^2 - K^2)}{15J(2J+1)(J-1)(2J-1)}$	$\frac{4(J^2 - K^2)((J-1)^2 - K^2)}{15J(2J+1)(J-1)(2J-1)}$

Tab. 9 Kleinsignalverstärkungskoeffizienten A_{JK} , B_{JK} und C_{JK} für R -, P - und Q -Zweig-Pumpen nach /63/.

Fälle in Tab. 9 angegeben. μ_{IR} ist das induzierte Dipolmoment, das mit der Anregung der ν_3 -Normalschwingung von CH_3F verbunden ist, μ_o das permanente Dipolmoment des Moleküls. Es gilt $\mu_{IR} = 0.205$ Debye, $\mu_o = 1.905$ Debye (1 Debye = $3.33 \cdot 10^{-30}$ Cb-m). $\frac{N}{V} \rho_{JK}^0$ gibt die Konzentration der Moleküle an, die sich im Anfangszustand $|0, J, K\rangle$ befinden.

Die abstimmbare Erzeugung von FIR-Strahlung erfolgte in dieser Arbeit durch R -Zweig-Pumpen. Welche FIR-Frequenzen dabei erzeugt werden können, soll im folgenden anhand von Abb. 25 und 26 diskutiert werden. Abb. 25 zeigt schematisch den doppelten Raman-prozeß für das R -Zweig-Pumpen für 3 verschiedene Pumpfrequenzen, wobei das Ausgangsniveau $|1\rangle$ mit $|0, J\rangle$ bezeichnet sei. (Wie schon in Abschnitt I.2.1.1 diskutiert, wird der dominante Beitrag zur Kleinsignalverstärkung durch Zustände mit $K = 3$ geliefert, da diese die größte relative Besetzungszahl aufweisen. Für die folgende Diskussion sei daher die Abhängigkeit der Verstärkung von K ignoriert; es sei stets $K = 3$ angenommen.) Daß der Energieabstand der Rotationszustände $|0, J-1\rangle$ und $|0, J\rangle$ geringer ist als der Abstand der Zustände $|1, J\rangle$ und $|1, J+1\rangle$, wie in Abb. 25 angedeutet ist, läßt sich mit Gl.(23) nachrechnen. Die durchgezogene Linie in Abb. 26 zeigt das Ergebnis einer numerischen Berechnung der Abhängigkeit der Kleinsignalverstärkung von der Frequenz ν_p der Pumpstrahlung nach Danly et al. /63/ für den Fall des Ausgangsniveaus $|0, J = 22\rangle$. (Hierin wurden auch die Beiträge der Zustände mit $K \neq 3$ berücksichtigt!) Das Ergebnis aus Abb. 26 läßt sich folgendermaßen verstehen:

Im Falle von $R(J-1) < R(J) < \nu_p$ (Abb. 25 a bzw. Punkt ① in Abb. 26) gilt $\delta p_1 = \nu_p - R(J) > 0$ und $\delta p_2 = \nu_p - R(J-1) > 0$. Die Abweichung δp_2 der Frequenz ν_p von der Resonanzfrequenz $R(J-1)$ ist jedoch viel größer als die entsprechende Abweichung δp_1 . Der wesentliche Beitrag zur Verstärkung $\alpha(\nu_p)$ wird daher durch den ersten Summanden in Gl.(32) geliefert; es gilt also $\alpha(\nu_p) \sim \frac{1}{\delta p_1^2}$. Eine Verringerung von ν_p in Richtung der Resonanzfrequenz $R(J)$ bewirkt somit eine Zunahme der Verstärkung.

Wird ν_p soweit erniedrigt, daß $R(J-1) < \nu_p < R(J)$ gilt (Abb. 25 b bzw. Punkt ② in

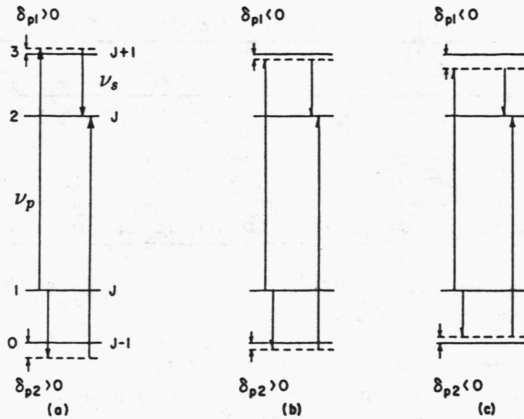


Abb. 25 Schematische Darstellung des doppelten Ramanprozesses für das R -Zweig-Pumpen für 3 verschiedene Pumpfrequenzen.

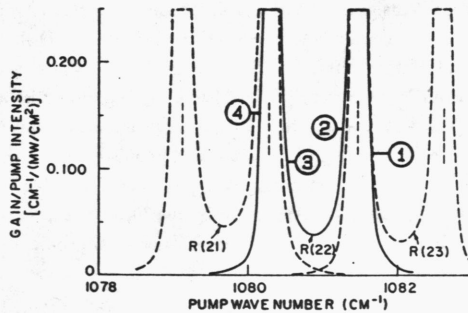


Abb. 26 Kleinsignalverstärkung für die stimulierte Ramanstreuung als Funktion der Frequenz ν_p der Pumpstrahlung (R -Zweig-Pumpen) nach /63/. Die gestrichelten senkrechten Linien geben die Lage der Resonanzfrequenzen $R(20)$, $R(21)$, $R(22)$ und $R(23)$ an. Die Erklärung der Verstärkungskurven findet sich im Text.

Abb. 26), ändert δp_1 sein Vorzeichen ($\delta p_1 < 0$), während nach wie vor $\delta p_2 > 0$ gilt. Der Interferenzterm in Gl.(32) wird positiv (Tab. 9 zeigt, daß $C_{JK} \geq 0 \forall J, K$). Die Verstärkung $\alpha(\nu_p)$ nimmt daher mit wachsendem Abstand von ν_p und $R(J)$ schwächer als $\sim \frac{1}{\delta p_1^2}$ ab. Sie ist somit bei negativen Werten von δp_1 größer als bei den entsprechenden positiven Werten und ist daher nicht mehr symmetrisch um $R(J)$, wie sie es bei einem einfachen Ramanprozeß wäre. Bei einer weiteren Reduktion von ν_p wird $\delta p_2 < |\delta p_1|$, so daß die Verstärkung wieder ansteigt (Punkt ③ in Abb. 26), da dann die Ramanemission im Grundzustand überwiegt.

Unterschreitet ν_p schließlich die Resonanzfrequenz $R(J-1)$, d.h. $\nu_p < R(J-1) < R(J)$ (Abb. 25 c bzw. Punkt ④ in Abb. 26), sind sowohl δp_1 als auch δp_2 negativ. Der dominante Beitrag zur Verstärkung wird jetzt, da $|\delta p_2| \ll |\delta p_1|$ gilt, durch den zweiten Summanden in Gl.(32) geliefert. Es gilt $\alpha(\nu_p) \sim \frac{1}{\delta p_2^2}$; die Verstärkung nimmt also mit zunehmendem Abstand von ν_p und $R(J-1)$ rasch ab.

Das Abschneiden der Verstärkungskurve $\alpha(\nu_p)$ im Bereich der Maxima bei den Resonanzfrequenzen $R(J)$ und $R(J-1)$ trägt der Tatsache Rechnung, daß Gl.(32) im Falle $\delta p_1 \rightarrow 0$ und $\delta p_2 \rightarrow 0$ nicht mehr gültig ist.

Die gestrichelten Kurven in Abb. 26 zeigen die Kleinsignalverstärkung $\alpha(\nu_p)$ für die stimulierte Ramanstreuung an den Übergängen, die von den Niveaus $|0, J-1 = 21\rangle$ bzw. $|0, J+1 = 23\rangle$ ausgehen. Für eine gegebene Frequenz ν_p erfolgt die stimulierte Ramanstreuung an dem Übergang, der für diese Frequenz die höchste Kleinsignalverstärkung zeigt. Diese Aussage ist von entscheidender Bedeutung für die Berechnung der Frequenz der gestreuten Strahlung. Für Frequenzen ν_p mit $R(J-1) < \nu_p < R(J)$ bestimmt die Ramanstreuung am Übergang $|0, J\rangle \rightarrow |1, J\rangle$ die Kleinsignalverstärkung $\alpha(\nu_p)$. Die Frequenz ν_s der gestreuten Strahlung errechnet sich gemäß dem Schema in Abb. 25 zu:

$$\nu_s = \nu_p - R(J) + \Delta(J) \quad \text{für } R(J-1) < \nu_p < R(J) \quad (33a)$$

Hierbei ist

$$h\Delta(J) = h(2J(B^1 - D_{JK}^1 K^2) - 4D_J^1 J^3) \simeq h2B^1(J+1) \quad (33b)$$

der Energieabstand der Molekülzustände $|1, J+1\rangle$ und $|1, J\rangle$. Es besteht somit in diesem Fall ein linearer Zusammenhang zwischen der Frequenz ν_p der Pumpstrahlung und der Frequenz ν_s der gestreuten Strahlung.

Wenn jedoch ν_p die Resonanzfrequenz $R(J)$ überschreitet, wird die stimulierte Ramanstreuung am Übergang $|0, J+1\rangle \rightarrow |1, J+1\rangle$ dominant. Nach Gl.(33) gilt dann für die Frequenz ν_s :

$$\nu_s = \nu_p - R(J+1) + \Delta(J+1) \quad R(J) < \nu_p < R(J+1)$$

Mit Hilfe von Gl.(23) und Tab.7 rechnet man leicht nach, daß gilt: $\Delta(J) - R(J) < \Delta(J+1) - R(J+1)$. Dies bedeutet, daß die Frequenz ν_s einen höheren Wert besitzt als die einfache Weiterführung von Gl.(33) über $R(J)$ hinaus liefern würde. Wenn die Frequenz ν_p der Pumpstrahlung also die Frequenz eines resonanten Übergangs $R(J)$ in Richtung höherer Frequenzen überschreitet, tritt eine Unstetigkeitsstelle in der Abstimmkurve $\nu_s = \nu_s(\nu_p)$ auf; die Frequenz ν_s nimmt schlagartig um den Betrag $2(B^0 - B^1)(J+1)$ zu (dieser Wert folgt aus den Gln.(23) und (33)).

Die vollständige Abstimmkurve $\nu_s = \nu_s(\nu_p)$, wenn ν_p kontinuierlich alle Frequenzen innerhalb des 9R-Emissionsbereichs des Hochdruck- CO_2 -Lasers durchläuft (Abb.16), ist in Abb.27 a gezeigt. Die senkrechten Striche kennzeichnen dabei die Position der resonanten Rotations-Schwingungs-Übergänge $R(J)$ in CH_3F , bei deren Überschreiten jeweils der diskutierte Sprung in der Frequenz ν_s der gestreuten Strahlung auftritt. Bei einer Variation der Frequenz des CO_2 -Lasers zwischen 1069 cm^{-1} und 1086 cm^{-1} erhält man somit FIR-Strahlung im Bereich von 23 cm^{-1} bis 45 cm^{-1} . Die tatsächlich emittierten Frequenzen überdecken dabei etwa 75% dieses Bereichs, wie man mit Hilfe der Gln.(23) und (33) leicht nachrechnet. Abb.27 b zeigt noch die Abstimmkurve, die man bei Verwendung des isotopischen $^{13}\text{CH}_3\text{F}$ erhalten würde. In diesem Fall müssen Emissionsfrequenzen des Hochdruck- CO_2 -Lasers im 9P-Emissionsbereich verwendet werden (1038 cm^{-1} - 1058 cm^{-1} ; Tab.8). Man erhält dann FIR-Strahlung im Bereich von 11 cm^{-1} bis 34 cm^{-1} . Experimente mit $^{13}\text{CH}_3\text{F}$ wurden jedoch in dieser Arbeit nicht durchgeführt.

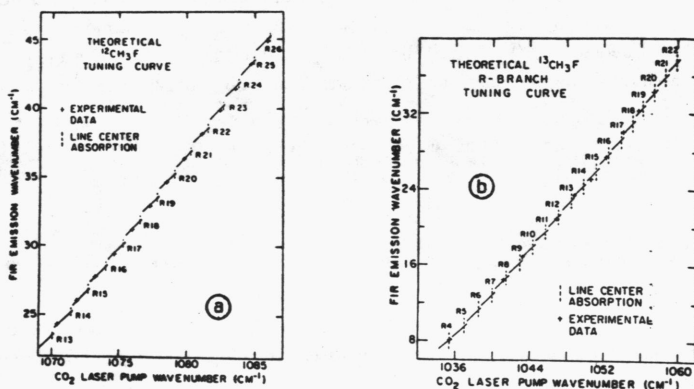


Abb.27 Abstimmkurven $\nu_s = \nu_s(\nu_p)$ für das R-Zweig-Pumpen nach /15/ in $^{12}\text{CH}_3\text{F}$ (a) und $^{13}\text{CH}_3\text{F}$ (b).

Wie bereits erwähnt, kann eine Erzeugung von FIR-Strahlung durch stimulierte Ramanstreuung in CH_3F auch durch *P*- bzw. *Q*-Zweig-Pumpen erfolgen. Eine ausführliche Diskussion des in diesen Fällen zu erwartenden Abstimmbereichs findet sich bei Danly et al. /63/. Beide Fälle sind jedoch in der Praxis von geringer Bedeutung, da beim *Q*-Zweig-Pumpen der tatsächliche Abstimmbereich sehr gering ist, während beim *P*-Zweig-Pumpen die Kleinsignalverstärkung für die stimulierte Ramanstreuung sehr viel geringer ist als beim *R*-Zweig-Pumpen /63/. Ursache ist, daß beim *P*-Zweig-Pumpen die Übergänge von Zuständen $|0, J\rangle$ mit $J \geq 32$ (Tab. 8) ausgehen, die eine sehr geringe relative Besetzung aufweisen (Abb. 23).

2.2 Aufbau des CH_3F -Raman-Lasers

Abb. 28 zeigt den vollständigen Aufbau des aus Hochdruck- CO_2 -Laser und CH_3F -Raman-Laser bestehenden Lasersystems zur abstimmbaren Erzeugung von FIR-Strahlung. Der CH_3F -Laser, der ähnlich dem von Mathieu et al. /16/ veröffentlichten Laser konstruiert wurde, bestand aus einem Wellenleiterrohr aus Quarzglas (Länge 1.2 m, Innendurchmesser 7 mm), das in einem Rohr aus V_2A -Stahl (Länge 1.25 m, Durchmesser 35 mm) montiert war. Als Eintrittsfenster für die CO_2 -Laserstrahlung wurde ein BaF_2 -Fenster verwendet, das Austrittsfenster für die FIR-Strahlung bestand aus einer 3 mm dicken TPX-Scheibe, die für Strahlung der Frequenz $\leq 40 \text{ cm}^{-1}$ eine Transmission von 85% besaß. Es wurde auch kristalliner Quarz als Material für das Austrittsfenster getestet, der für Strahlung aus dem 9R-Emissionsbereich des Hochdruck- CO_2 -Lasers ein Reflexionsvermögen von etwa 70% besitzt /16/ und somit eine bessere Ausnutzung der Pumpstrahlung erlaubt hätte. Die Absorption der restlichen nichtreflektierten Strahlung im Quarzfenster führte jedoch zu allmählichen Oberflächenschäden, insbesondere wenn die Ausbreitungsrichtung der CO_2 -Laserstrahlung von der optischen Achse abwich und so eine Fokussierung der Strahlung durch die Krümmung des Wellenleiterrohres an dessen Ende erfolgte. Dieses Problem trat bei TPX nicht auf. Da die nach dem Durchlauf durch die Ramanröhre noch übrigbleibende Strahlung des CO_2 -Lasers jedoch nicht vollständig vom TPX-Fenster absorbiert wurde, wurde sie außerhalb der Röhre durch eine Platte aus kristallinem Quarz abgeblockt. Sowohl das Eintritts- wie auch das Austrittsfenster waren mittels Federbälgen aus Edelstahl am Rohr des Raman-Lasers befestigt. Damit war eine Einstufung der Fenster senkrecht zur optischen Achse möglich, um so die Rückkopplung für die FIR-Strahlung - sowohl BaF_2 wie TPX reflektieren einen geringen Bruchteil dieser Strahlung - zu optimieren.

Der Gasdruck im Inneren der Ramanzelle wurde mit Hilfe eines Kapselfeder-Vakuummeters (Fa. Leybold, 0-100 Torr) kontrolliert. Die typischen Drücke lagen im Bereich von 10-30 Torr. Mit Hilfe einer Kryopumpe war es möglich, das relativ teure CH_3F -Gas in einen speziellen Aufbewahrungsbehälter zu pumpen. Auf diese Weise ging kein Gas verloren, wenn Druckreduktionen erforderlich waren.

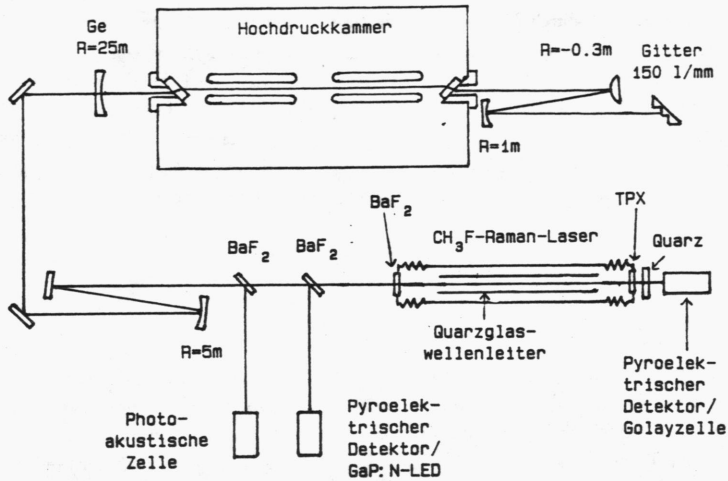


Abb. 28 Optische Anordnung des Hochdruck-CO₂-Lasers und CH₃F-Raman-Lasers.

Die Strahlung des CO₂-Lasers wurde mit Hilfe eines konkaven Molybdänspiegels auf den Eingang des Wellenleiterrohres fokussiert. Dort koppelt die Strahlung des Pumplasers, die, wie in Abschnitt I.1.3.5 diskutiert, eine annähernd Gaußsche Querschnittsverteilung wie eine sich frei ausbreitende TEM₀₀-Mode besitzt, an die Moden des Quarzglaswellenleiters. Nach Hall et al. /83/ erhält man eine maximale Kopplung, nämlich 98%, zwischen einer sich frei ausbreitenden TEM₀₀-Mode und der EH₁₁-Wellenleitermode, die die geringsten Dämpfungsverluste in einem Quarzglaswellenleiter besitzt /63/, wenn für die Brennweite des fokussierenden Hohlspiegels gilt:

$$f = 0.644 \pi \frac{a D}{2 \lambda} \quad (34)$$

Hierbei ist a der Radius des Wellenleiterrohres, D die $\frac{1}{\sqrt{2}}$ -Breite für die Querschnittsverteilung der Intensität der Pumpstrahlung. Im vorliegenden Fall galt $a = 3.5$ mm, $D \approx 8$ mm und $\lambda \approx 9.3$ μ m. Die optimale Brennweite betrug $f \approx 3$ m. Da ein Spiegel mit solcher Brennweite nicht zur Verfügung stand, wurde ein Spiegel mit $f = 2.5$ m verwendet. Die CO₂-Strahlung hatte dann einen Durchmesser ($\frac{1}{\sqrt{2}}$ -Breite) von etwa 4 mm an der Eintrittsöffnung des Wellenleiters. Die Transmission des leeren Wellenleiters betrug 85%. Dies war einerseits eine Folge der nicht ganz optimalen Einkopplung; andererseits bewirken geringfügige

Krümmungen des Wellenleiterrohres in Längsrichtung, wie sie durch die Montage im Innern der Stahlröhre stets vorkamen, eine deutliche Erhöhung der Dämpfungsverluste /83/.

Wie in Abschnitt I.1.3.5 geschildert, wurde ein Teil der Strahlung des CO₂-Lasers mittels einer BaF₂-Scheibe in eine photoakustische Zelle mit CH₃F-Gas eingekoppelt, um die Emissionsfrequenz des Hochdruck-CO₂-Lasers überprüfen zu können. Mit einer zweiten BaF₂-Scheibe wurde ein weiterer Teil der Pumpstrahlung ausgekoppelt und auf einen Pyrodetektor (Gentec ED100, Ansprechempfindlichkeit 134 V/J) fokussiert, um die Pulsenergie des CO₂-Lasers zu kontrollieren. Der Pyrodetektor konnte wieder durch eine als Photoleitungs-detektor wirkende, auf 4.2 K gekühlte GaP:N-Leuchtdiode ersetzt werden, falls die zeitliche Struktur der in den CH₃F-Raman-Laser eintretenden Strahlung erfaßt werden sollte. Die Beobachtung der zeitlichen Struktur der FIR-Laserpulse erfolgte mit Hilfe eines schnellen Pyrodetektors (Molelectron P5-01, Anstiegszeit 0.5 nsec) bzw. mit Hilfe von Schottkydioden (SD017 der Fa. Farran Technology, Sperrschichtkapazität 2.5 fF). Meßtechnische Aspekte dazu werden in Abschnitt I.3.2 behandelt. Zur Messung der Pulsenergie der FIR-Strahlung wurden eine Golayzelle mit Diamantfenster bzw. der pyroelektrische Detektor P5-01 in der Energiemeßmode verwendet. Die Golayzelle wurde nur für eine relative Messung der Pulsenergie benutzt, eine Abschätzung des Absolutwertes der Pulsenergie erfolgte mit Hilfe des Pyrodetektors. Das Verfahren dabei war wie folgt: Im Abstand von 10 cm vom Ende des Wellenleiterrohres ließ sich die Querschnittsverteilung der Energiedichte der FIR-Strahlung als eine annähernd Gaußsche Verteilung mit einer Halbwertsbreite (FWHM) von $\Delta = 4.5$ mm beschreiben. Unter Energiedichte sei im folgenden der durch eine Flächeneinheit tretende Teil der gesamten Pulsenergie verstanden. Der Pyrodetektor mit einem kreisförmigen Detektorelement (Durchmesser $D_p = 1$ mm) wurde im Zentrum der Verteilung angebracht. Für die Energiedichte \tilde{E}_0 der Laserstrahlung in diesem Zentrum galt dann, da $D_p \ll \Delta$ war:

$$\tilde{E}_0 = \frac{4}{D_p^2 \pi} \frac{U}{R_o} \quad (35)$$

Hierbei war U die vom Pyrodetektor gelieferte Spitzenspannung und R_o dessen Ansprechempfindlichkeit in $\frac{V}{J}$. Die Gesamtenergie E_0 des Laserpulses war dann

$$E_0 = \int_0^\infty \int_0^\infty \tilde{E}_0 \exp\left(-\frac{4 \ln 2}{\Delta^2}(x^2 + y^2)\right) dx dy = \frac{1}{\ln 2} \frac{\Delta^2}{D_p^2} \frac{U}{R_o} \quad (36)$$

Die Ansprechempfindlichkeit R_o des Pyrodetektors lag nach Herstellerangaben bei voller Absorption der auffallenden Strahlung in der Oberflächenbeschichtung des Detektorkristalls bei $5.7 \frac{mV}{\mu J}$ für eine kapazitive Last von 70 pF. (Diese war durch Kabel und Oszillograph gegeben.) Da jedoch die Beschichtung des Detektors für FIR-Strahlung ein endliches Reflexionsvermögen besaß, wirkte dieser insbesondere wegen seines hohen Brechungsindex

als Fabry-Perot-Etalon für die FIR-Strahlung. Die Reflexion des Detektorkristalls war somit wellenlängenabhängig und konnte Werte von über 80% erreichen /84/. Die tatsächliche Ansprechempfindlichkeit des Detektors ist somit eine Funktion der Frequenz ν_{FIR} der FIR-Strahlung, d.h. $R = R(\nu_{FIR})$, wobei stets gilt $R(\nu_{FIR}) \leq R_0$. Beispielsweise ergab ein Vergleich der Signale des Pyrodetektors bei einer Frequenz $\nu_{FIR} = 29.35 \text{ cm}^{-1}$ mit denen der Golayzelle, deren Ansprechempfindlichkeit - sie kann nach Herstellerangaben als unabhängig von der Frequenz der einfallenden Strahlung angenommen werden - mit Hilfe der $9.3 \text{ }\mu\text{m}$ -Strahlung des Hochdruck- CO_2 -Lasers abgeschätzt wurde, eine Ansprechempfindlichkeit des Pyrodetektors von $R(29.35 \text{ cm}^{-1}) \simeq 0.2 \dots 0.5 R_0$. Die Gesamtenergie E_0 der FIR-Laserpulse bei einer Frequenz $\nu_{FIR} = 29.35 \text{ cm}^{-1}$ ergab sich also aus Gl.(36) unter Einfügen eines multiplikativen Korrekturfaktors γ , dessen Wert zwischen 2 und 5 liegt. Durch die Auswertung der Experimente zur nichtlinearen Absorption von FIR-Strahlung in Sauerstoff-dotiertem Silizium, von denen in Kapitel II dieser Arbeit berichtet wird, läßt sich der Korrekturfaktor sogar auf einen Wert $\gamma \simeq 3.5$ festlegen (siehe Abschnitt II.3). (Eine Verwendung der Golayzelle als frequenzunabhängiger Detektor zur Bestimmung des Absolutwerts der Pulsenergie war nicht möglich, da sie für andere Zwecke benötigt wurde - siehe Abschnitt I.3.1.)

Die Bestimmung der Frequenz der emittierten FIR-Strahlung erfolgte mit Hilfe eines Fabry-Perot-Interferometers. Dieses bestand aus 2 auf Ringe aufgespannte metallische Kreuzgitter (Gitterkonstante $35 \text{ }\mu\text{m}$, Drahtstärke $7 \text{ }\mu\text{m}$). Eine Änderung des Abstands der beiden Gitter war mit einem schrittmotorgetriebenen Präzisionsschlitten, auf den eines der Gitter montiert war, in Einheiten von $1 \text{ }\mu\text{m}$ möglich. Detaillierte Informationen über solche Interferometer finden sich in /85,86/. Mit Hilfe der $336 \text{ }\mu\text{m}$ -Strahlung eines HCN-Lasers (Bandbreite $\sim 10 \text{ MHz}$ /87/) wurde die Finesse \mathcal{F} des Interferometers zu $\mathcal{F} = 45$ bestimmt; dies entspricht einem Reflexionsvermögen der Kreuzgitter von 93%. In der 60. Ordnung lag damit das Auflösungsvermögen bei $\frac{\Delta\nu}{\nu} = 3.7 \cdot 10^{-4}$. Die maximale Transmission betrug 30%; als Detektor der durch das Fabry-Perot-Interferometer transmittierten Strahlung wurde die Golayzelle verwendet. Die Aufnahme eines Interferogramms erfolgte schrittweise in Abständen von $1 \text{ }\mu\text{m}$; der erwähnte ECB-Bus-Rechner steuerte den Schrittmotor. Das Verfahren wird in Abschnitt I.3.2 genauer geschildert; ein Beispiel eines Interferogramms findet sich im folgenden Abschnitt. Als Test wurde die resonante $496.1 \text{ }\mu\text{m}$ -Laserlinie von CH_3F vermessen, die man durch Pumpen mit der $9P(20)$ -Linie eines CO_2 -Lasers erhält. Das Resultat $(496 \pm 0.3) \text{ }\mu\text{m}$ stimmt hervorragend mit dem tatsächlichen Wert überein.

2.3 Eigenschaften des CH_3F -Raman-Lasers

Die Erzeugung der FIR-Strahlung erfolgte durch R -Zweig-Pumpen von $^{12}\text{CH}_3\text{F}$ mit Hilfe der $9R$ -Emissionsstrahlung des Hochdruck- CO_2 -Lasers, wie in Abschnitt I.2.1.2 diskutiert. Die Resonatorkonfiguration des letzteren war die in Abb. 15 gezeigte.

2.3.1 Emissionscharakteristik

Abb. 29 zeigt die vollständige Emissionscharakteristik des CH_3F -Raman-Lasers, d.h. die Pulsenergie der emittierten Strahlung als Funktion ihrer Frequenz für einen Gasdruck von 17 Torr. Die Frequenz des CO_2 -Lasers wurde hierbei von 1069 cm^{-1} bis 1085 cm^{-1} ($9R(6) - 9R(30)$) in Schritten von 0.04 cm^{-1} variiert; seine Pulsenergie war konstant auf 180 mJ gehalten worden. Die Frequenz der FIR-Strahlung wurde nach Gl.(33) aus der Frequenz der CO_2 -Laserstrahlung berechnet, die wiederum gemäß der Diskussion in Abschnitt I.1.3.5 mit Hilfe der Winkelposition des Gitters und der photoakustischen Zelle ermittelt wurde. Als Detektor wurde die Golayzelle wegen der Unabhängigkeit ihrer Ansprechempfindlichkeit von der Frequenz verwendet. Da diese jedoch eine sehr hohe Empfindlichkeit besitzt, mußte die FIR-Strahlung über einen konvexen Spiegel aufgeweitet werden, um eine Überlastung zu vermeiden. Der Einsatz von Abschwächern verbot sich, da deren Abschwächungsfaktor durch Vielstrahlinterferenzen innerhalb und zwischen den Abschwächern frequenzabhängig war (siehe Abschnitt I.3.3). Mit Hilfe der Golayzelle konnte somit nur eine Relativmessung der Pulsenergie erfolgen. Eine absolute Energiemessung erfolgte bei einer FIR-Frequenz von 29.35 cm^{-1} mit Hilfe des Pyrodetektors gemäß dem in I.2.2 beschriebenen Verfahren (Korrekturfaktor $\gamma = 3.5$).

Das Emissionsspektrum in Abb. 29 zeigt, daß eine kontinuierliche Erzeugung von FIR-Strahlung im Frequenzbereich von 23 cm^{-1} bis 45 cm^{-1} in einzelnen Intervallen möglich ist, die eine typische Breite von 1.2 cm^{-1} und untereinander einen Abstand von etwa 0.4 cm^{-1} besitzen. Der tatsächliche Abstimmbereich überdeckt somit 70% des Bereichs von 23 cm^{-1} bis 45 cm^{-1} . Dies ist in ausgezeichneter Übereinstimmung mit der theoretischen Abstimmkurve, die zum Vergleich in Abb. 29 b miteingezeichnet ist.

Innerhalb der einzelnen Intervalle zeigen sich starke Variationen der FIR-Pulsenergie. So wird das scharfe Minimum, das etwa im Zentrum eines jeden Abstimmintervalls liegt, durch einen Zwei-Photonen-Absorptionsprozeß verursacht. Liegt nämlich die Frequenz ν_p des Pumplasers in der Mitte zwischen den Frequenzen von 2 benachbarten Rotations-Schwingungs-Übergängen in CH_3F , d.h. $\nu_p = \frac{1}{2}(R(J) + R(J-1))$, entspricht die Summe aus der Frequenz des CO_2 -Photons und der Frequenz $\nu_s = \nu_s(\nu_p)$ des zugehörigen FIR-Photons, wie man mittels der Gln.(23) und (33) nachrechnet, gerade der Frequenz des Übergangs $|0, J-1\rangle \rightarrow |1, J+1\rangle$. Die Position dieser Zwei-Photonen-Absorption bietet eine zusätzliche Kontrollmöglichkeit für die Emissionsfrequenz des CO_2 -Lasers. Daß die Pulsenergie der FIR-Strahlung für Frequenzen, die an den Rändern eines jeden Abstimmbereichs liegen und die damit Frequenzen des Pumplasers in der Nähe der Resonanzfrequenzen $R(J)$ entsprechen, abnimmt, widerspricht der theoretischen Berechnung der Kleinsignalverstärkung, die ja gerade für Frequenzen ν_p in der Nähe von $R(J)$ eine starke Zunahme der Kleinsignalverstärkung voraussagt (Abb. 26). Die Ursache dieses Phänomens wird in Abschnitt I.2.3.2 untersucht.

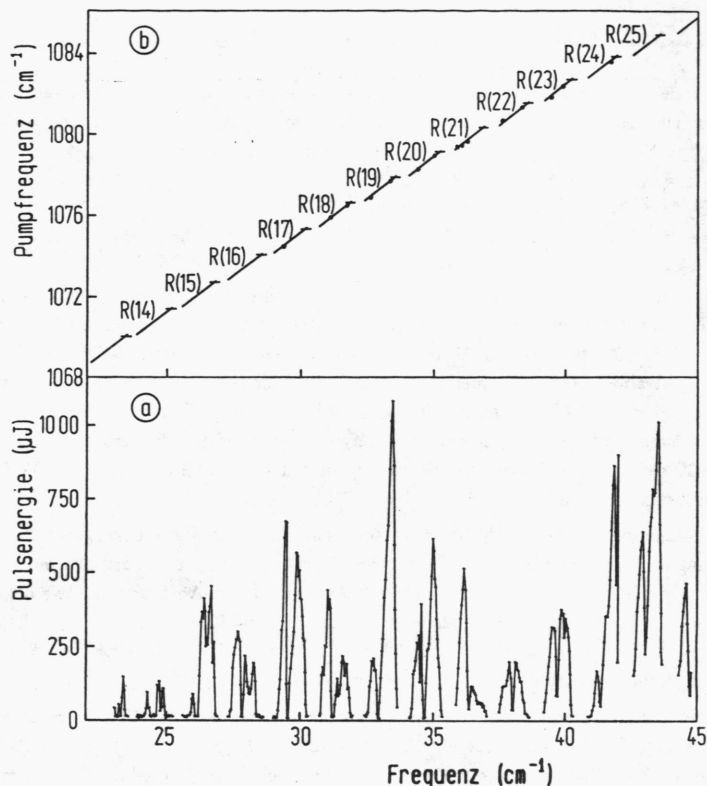


Abb. 29 Emissionscharakteristik des CH_3F -Raman-Lasers (a) mit theoretischer Abstimmkurve (b).

Die waagrechten Striche in (b) geben die Positionen der Rotations-Schwingungs-Übergänge $R(J)$ an, die Punkte zeigen Kontrollmessungen der Abstimmcharakteristik mit Hilfe eines Fabry-Perot-Interferometers.

Es sei noch erwähnt, daß wiederholte Messungen der Emissionscharakteristik des CH_3F -Raman-Lasers bei gleicher Energie des Pumplasers zwar stets in den wesentlichen Strukturen übereinstimmen, daß aber die relative Höhe der einzelnen Maxima der verschiedenen Abstimmintervalle von der Justierung des Ein- bzw. Austrittsfensters der Ramanzelle abhing.

Dies zeigt, daß die in Abschnitt I.2.2 erwähnte Rückkopplung der FIR-Strahlung bei der stimulierten Ramanemission eine Rolle spielt.

Eine Kontrolle der FIR-Frequenz wurde an vielen Stellen innerhalb der in Abb. 29 gezeigten Emissionscharakteristik mit Hilfe des im vorhergehenden Abschnitt erwähnten Fabry-Perot-Interferometers durchgeführt. Abb. 30 a zeigt das Beispiel eines Interferogramms für eine FIR-Emissionsfrequenz von 34.75 cm^{-1} , die nach der am Anfang dieses Kapitels beschriebenen Methode eingestellt worden war. Die Auswertung des Interferogramms (Maxima A) liefert für die Wellenlänge der emittierten Strahlung einen Wert von $287.3 \text{ }\mu\text{m}$, was einer Frequenz von 34.8 cm^{-1} entspricht. Man erhält also eine sehr gute Übereinstimmung. Der Fehler von 0.05 cm^{-1} resultiert zum einen daraus, daß die Emissionsfrequenz des Hochdruck- CO_2 -Lasers nur auf $\pm 0.04 \text{ cm}^{-1}$ eingestellt werden konnte (dies entspricht einem Winkelschritt des schrittmotorgetriebenen Gitterdrehtisches - siehe Abschnitt I.1.3.5); zum anderen konnte auch die Abstandsänderung des Fabry-Perot-Interferometers, die für die Aufnahme der 4 aufeinanderfolgenden Maxima in Abb. 30 a nötig war, nur auf $\pm 1 \text{ }\mu\text{m}$ (entsprechend der kleinstmöglichen Verschiebung des Präzisionsschlittens des Interferometers) genau angegeben werden, was einem Fehler von $\pm 0.3 \text{ }\mu\text{m}$ bzw. $\pm 0.04 \text{ cm}^{-1}$ entspricht. Als Ergebnis läßt sich daher festhalten, daß die Einstellgenauigkeit der FIR-Frequenz über die Emissionsfrequenz des Hochdruck- CO_2 -Lasers etwa $\pm 0.05 \text{ cm}^{-1}$ beträgt. Die Punkte in Abb. 30 b geben die Resultate der analog dem obigen Beispiel durchgeführten Kontrollmessungen verschiedener FIR-Frequenzen an.

Mit Hilfe des Interferometers wurde auch die spektrale Verteilung der emittierten FIR-Strahlung untersucht. Das Interferometer wurde dabei in 60. Ordnung betrieben und besaß somit für Strahlung der Frequenz 34.8 cm^{-1} eine Auflösung von 0.013 cm^{-1} (390 MHz) (Abschnitt I.2.2). Die Auswertung der Interferogramme zeigte, daß sich die spektrale Verteilung der emittierten Strahlung ($\nu_{\text{FIR}} = 34.8 \text{ cm}^{-1}$) durch eine Gaußsche Verteilung beschreiben ließ, deren Halbwertsbreite (FWHM) etwa $\Delta\nu_L = (4 \pm 0.5) \text{ GHz}$ betrug und somit vergleichbar der Breite der Strahlung des Hochdruck- CO_2 -Lasers (4.8 GHz) war.

Das Interferogramm in Abb. 30 a zeigt, daß dort neben der (durch die Emissionsfrequenz des Hochdruck- CO_2 -Lasers eingestellten) Frequenz von 34.8 cm^{-1} (Maxima A) noch eine weitere Komponente mit der Frequenz 33.5 cm^{-1} enthalten war (Maxima B). Noch deutlicher ist dies in dem Interferogramm von Abb. 30 b zu sehen, das für eine Einstellung der Emissionsfrequenz des FIR-Lasers auf 36.47 cm^{-1} aufgenommen worden war. Hier beinhaltete die emittierte Strahlung 3 Frequenzkomponenten, nämlich die eingestellte Frequenz von 36.5 cm^{-1} (Maxima A) sowie die Frequenzen 35.2 cm^{-1} und 33.5 cm^{-1} (Maxima B und C). Der Ursprung dieser zusätzlichen Frequenzkomponenten ist den Energieniveauschemata in Abb. 30 zu entnehmen. Es handelt sich um der Ramanemission nachfolgende resonante reine Rotationsübergänge im angeregten Schwingungszustand $v = 1$. Geht ein CH_3F -Molekül beispielsweise bei der Ramanstreuung eines CO_2 -Photons vom Zustand $|0, J\rangle$ in den Zustand $|1, J\rangle$ über, so ist letzterer gegenüber dem Zustand $|1, J-1\rangle$ überbesetzt. Es

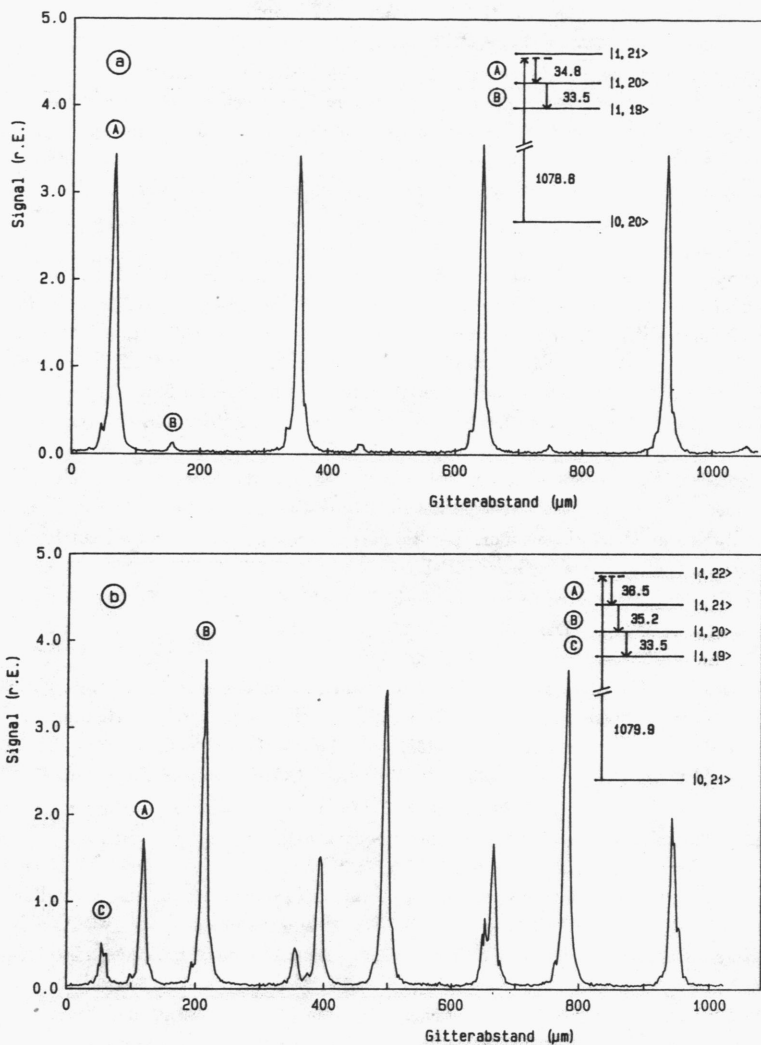


Abb. 30 Fabry-Perot-Interferogramme der emittierten FIR-Strahlung bei einer Frequenzeinstellung von 34.75 cm^{-1} (a) und 36.47 cm^{-1} (b). (Die eingezeichneten Niveauschemata werden im Text diskutiert.)

kommt zur stimulierten Emission von Strahlung der Frequenz, die dem Energieabstand der Zustände $|1, J\rangle$ und $|1, J-1\rangle$ entspricht. Dieser Prozeß setzt sich kaskadenartig zu niedrigeren Rotationszuständen fort. (Solche Übergänge wurden jüngst auch von Izatt et al. /88/ beobachtet.)

Diese Kaskadenübergänge traten nur bei sehr hohen FIR-Pulsenergien ($> 250 \mu\text{J}$) auf; denn der Aufbau einer Überbesetzung des Zustands $|1, J\rangle$ gegenüber dem Zustand $|1, J-1\rangle$ konkurriert mit dem R-T-Transfer, durch den sich in einer Zeit $\tau_{RT} \simeq 0.5 \text{ nsec}$ (CH_3F -Druck 17 Torr; siehe Abschnitt I.2.1.1) eine der Gasttemperatur entsprechende Besetzungsverteilung der Rotationsniveaus im angeregten Schwingungszustand einstellt. Die Pulsenergie betrug etwa $350 \mu\text{J}$ im Falle von Abb. 30 a. Der Energieanteil der zweiten Frequenzkomponente betrug etwa 3% und konnte somit vernachlässigt werden. Dennoch konnten Situationen auftreten, wo dies nicht mehr möglich war. Ein Beispiel ist der in Abb. 30 b gezeigte Fall: Dort wurde die Ramanstrahlung (Frequenz 36.5 cm^{-1}) auf dem Weg zum Detektor durch Wasserdampf in Luft, der bei 36.6 cm^{-1} eine starke Absorptionslinie hat /89/, so stark geschwächt, daß der Detektor im wesentlichen nur mehr die Strahlung der Kaskadenübergänge, die nicht vom Wasserdampf absorbiert wurde, registrierte. Wäre es der Sinn des Experiments gewesen, den Absorptionskoeffizienten für Strahlung der Frequenz 36.5 cm^{-1} in Luft zu bestimmen, hätte das Experiment falsche Resultate geliefert. Eine sorgfältige Analyse der spektralen Verteilung der am Detektor ankommenden Strahlung ist also angebracht! Ein weiteres Beispiel dafür findet sich in Kapitel II dieser Arbeit.

2.3.2 Wirkungsgrad

Die Untersuchungen zum Wirkungsgrad des CH_3F -Raman-Lasers erfolgten für eine Emissionsfrequenz von 35.1 cm^{-1} . Die zugehörige Frequenz des Hochdruck- CO_2 -Lasers ($\nu_p = 1078.9 \text{ cm}^{-1}$) lag knapp (0.2 cm^{-1}) unterhalb der Frequenz des Rotations-Schwingungs-Übergangs $R(20) = 1079.1 \text{ cm}^{-1}$. Sämtliche Intensitätsangaben der CO_2 -Laserstrahlung in diesem Abschnitt beziehen sich auf die Intensität am Anfang bzw. Ende des Quarzglaswellenleiterrohres in der CH_3F -Ramanzelle. Der Durchmesser des CO_2 -Laserstrahls betrug dort etwa 4 mm (siehe Abschnitt I.2.2).

Abb. 31 zeigt zunächst die zeitliche Korrelation zwischen den CO_2 -Laserpulsen und den zugehörigen FIR-Pulsen. Beide Pulse wurden simultan mit dem in Abschnitt I.3.2 beschriebenen Verfahren aufgezeichnet. Die Struktur der CO_2 -Laserpulse ist geprägt durch die in Abschnitt I.1.4.2 besprochene Kopplung der longitudinalen Moden, die innerhalb der Bandbreite der CO_2 -Laserstrahlung enthalten sind. Der zeitliche Abstand der einzelnen Spitzen entspricht wieder der Umlaufzeit der Photonen im Resonator ($L = 3.3 \text{ m}$). Die FIR-Pulse zeigen bei schwachen Pumpintensitäten die gleiche zeitliche Struktur wie die CO_2 -Laserpulse

(Abb. 31 a). Aus Abb. 31 a liest man auch ab, daß stimulierte Ramanemission erst ab einer Schwellenintensität $I_{th} \approx 4 \frac{\text{MW}}{\text{cm}^2}$ der CO_2 -Laserstrahlung eintritt. Dieser Wert ist vergleichbar einem aus Abb. 26 abschätzbaren Wert: Oft wird nämlich die Schwellenintensität über die Beziehung $\alpha(\nu_p) L \approx 30$ (L = Länge der Ramanzelle) definiert. Bei dieser Bedingung erreicht die Strahlung, die am Beginn der Ramanzelle durch spontane Ramanstreuung entsteht und dort typische Intensitätswerte von $10^{-12} - 10^{-13} \frac{\text{W}}{\text{cm}^2}$ besitzt, nach Durchlauf durch die Zelle experimentell detektierbare Intensitätswerte in der Größenordnung $1 - 10 \frac{\text{W}}{\text{cm}^2} / 63$. Für $\nu_p = 1078.9 \text{ cm}^{-1}$ folgt aus Abb. 26 eine Schwellenintensität $I_{th} = 3.3 \frac{\text{MW}}{\text{cm}^2}$. Infolge der endlichen Bandbreite des Meßwerterfassungssystems (siehe Abschnitt I.3.2) ist die Modulationstiefe der CO_2 -Laserpulse in Wirklichkeit stärker als in Abb. 31 gezeigt. Die Intensität der Pumplaserstrahlung unterschreitet somit nach jeder Spitze die Schwellenintensität I_{th} ; die FIR-Leistung geht dann, wie in Abb. 31 a unten gezeigt, auf Null zurück. Auch die Halbwertsbreite der im FIR-Puls enthaltenen Spitzen ist aus Gründen der Bandbreite der Meßelektronik schmaler als in Abb. 31 a. Die gesamte Dauer des FIR-Pulses beträgt etwa 100 nsec und entspricht somit der Zeitdauer, in der die Intensität des CO_2 -Laserpulses oberhalb der Schwellenintensität I_{th} liegt.

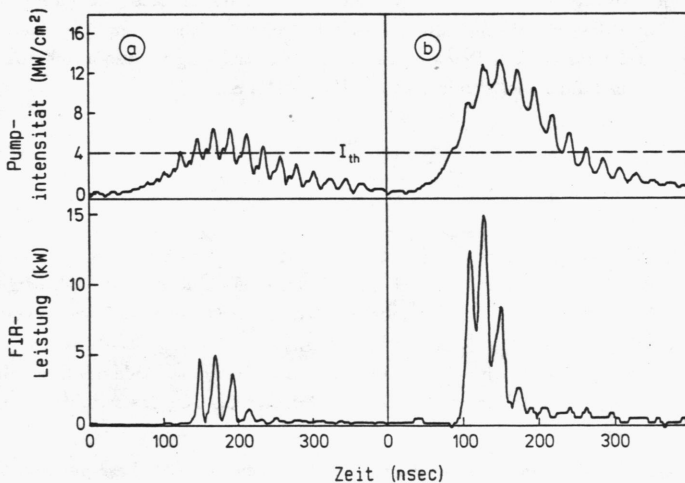


Abb. 31 CO_2 -Laserpulse (obere Kurven) und zugehörige FIR-Pulse (untere Kurven) für CO_2 -Laserpulsenergien von 100 mJ (a) und 250 mJ (b).

Bei hohen Intensitäten der CO_2 -Laserstrahlung erreicht die Leistung der FIR-Strahlung ihr

Maximum im ansteigenden Teil der CO₂-Laserpulse und nimmt bereits wieder ab, noch bevor die Intensität der Pumpstrahlung ihren Maximalwert angenommen hat. Insbesondere ist die Leistung der FIR-Strahlung bereits zu einem Zeitpunkt nahezu verschwunden, an dem die Intensität der CO₂-Laserstrahlung noch deutlich über der Schwellenintensität I_{th} liegt. Dies zeigt, daß die Kleinsignalverstärkung und damit der Wirkungsgrad für die Umsetzung der Photonen des CO₂-Lasers (Frequenz $\nu_p = 1078.9 \text{ cm}^{-1}$) in FIR-Photonen (Frequenz $\nu_{FIR} = 35.1 \text{ cm}^{-1}$) zeitabhängig ist.

Um die Ursache dieser zeitlichen Änderung der Kleinsignalverstärkung während des CO₂-Laserpulses zu ermitteln, wurde die Transmission der CH₃F-Ramanzelle für die Strahlung des CO₂-Lasers ($\nu_p = 1078.9 \text{ cm}^{-1}$) zeitaufgelöst untersucht. Dazu wurde die zeitliche Struktur der Laserpulse mit dem in Abschnitt I.3.2 beschriebenen Verfahren simultan vor und nach der Ramanzelle beobachtet. (Das Austrittsfenster aus TPX wurde in diesem Fall durch ein BaF₂-Fenster ersetzt; die Detektion der transmittierten Strahlung erfolgte wie die der einfallenden Strahlung mit einer auf 4.2 K gekühlten GaP:N-Leuchtdiode.)

Als Resultat ergab sich: Lag das Intensitätsmaximum der CO₂-Laserpulse unter der Schwellenintensität I_{th} für die stimulierte Ramanstreuung, waren die Pulsformen für die einfallende und transmittierte CO₂-Strahlung identisch; ein Teil der einfallenden Strahlung wurde jedoch im Innern der Ramanzelle absorbiert. Ein Vergleich der Transmission der Zelle mit und ohne CH₃F-Gas ergab, daß im Falle des verwendeten CH₃F-Gasdrucks von 21 Torr etwa 70% der einfallenden Strahlung absorbiert wurden. Dies entspricht einem Absorptionskoeffizienten $\alpha (1078.9 \text{ cm}^{-1}) = 0.01 \text{ cm}^{-1}$. Ein Blick auf die Linienbreite des Rotations-Schwingungs-Übergangs $R(20)$, die nach Abschnitt I.2.1.1 bei einem Druck von 21 Torr etwa 0.025 cm^{-1} beträgt, läßt den Schluß zu, daß die beobachtete Absorption der CO₂-Strahlung der Frequenz $\nu_p = 1078.9 \text{ cm}^{-1}$ der resonanten Absorption in den Flanken der Linie $R(20)$, deren Zentrum bei 1079.1 cm^{-1} liegt, zugeordnet werden kann, insbesondere da die CO₂-Strahlung eine Bandbreite von 0.16 cm^{-1} besitzt.

Bei hohen Intensitäten der CO₂-Laserpulse zeigte der transmittierte Puls eine deutlich andere zeitliche Form als der einfallende Puls (Abb. 32 a). So besaß er ein eindeutig erkennbares Minimum (Pfeil in Abb. 32 a) zu einem Zeitpunkt, an dem die Intensität des einfallenden Pulses noch zunahm. Das Minimum fällt zeitlich gesehen nahezu mit dem Emissionsmaximum der FIR-Strahlung (siehe Abb. 31 b) zusammen. Abb. 32 a erlaubt damit den Schluß, daß ein merklicher Anteil der einfallenden CO₂-Laser-Photonen durch stimulierte Ramanstreuung in FIR-Photonen umgewandelt wird. In Abb. 32 b ist die zeitliche Änderung der Transmission der Ramanzelle, die sich aus dem Verhältnis der Intensitäten der transmittierten und der einfallenden Strahlung ergibt, aufgetragen. Zu Beginn des CO₂-Laserpulses wird die Transmission ($\sim 35\%$) durch die erwähnte Resonanzabsorption der Laserstrahlung in der Flanke der $R(20)$ -Absorptionslinie bestimmt. Die Transmission durchläuft dann ein Minimum, das ein Zeichen für die starke Ramanstreuung der CO₂-Strahlung ist, und steigt anschließend

innerhalb von 50 nsec auf einen Wert von 60% an, und ist damit zum Zeitpunkt, an dem die Intensität des einfallenden Pulses ihr Maximum erreicht, deutlich höher als zu Beginn des Pulses. Dies zeigt, daß zu diesem Zeitpunkt ein großer Teil der CH_3F -Moleküle durch die Ramanstreuung bzw. die Resonanzabsorption der CO_2 -Strahlung vom Schwingungsgrundzustand in den angeregten Schwingungszustand überführt worden ist. Da die Lebensdauer des angeregten Schwingungszustands länger ist als die Dauer des CO_2 -Laserpulses (für 21 Torr beträgt $\tau_{\text{VV}} = 240$ nsec, siehe Abschnitt I.2.1.1), stehen Moleküle, die sich im angeregten Zustand befinden, für Ramanprozesse im restlichen Teil des CO_2 -Laserpulses nicht mehr zur Verfügung. Die Ursache der Zeitabhängigkeit der Kleinsignalverstärkung bzw. des Wirkungsgrades ist also die Entvölkerung des Schwingungsgrundzustandes durch die genannten Prozesse.

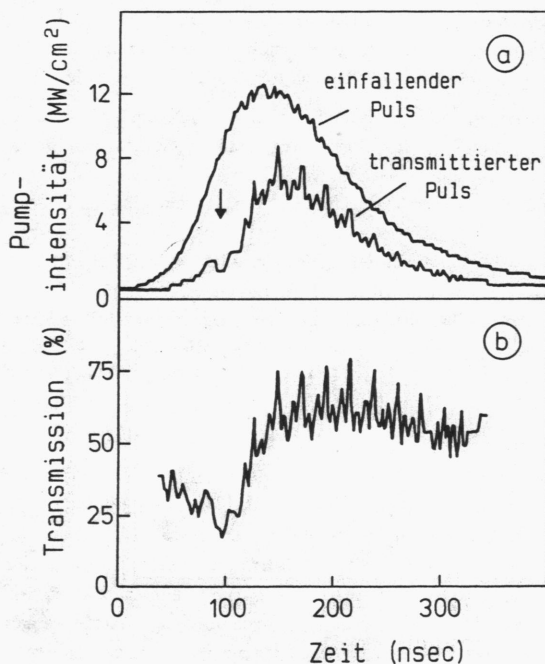


Abb. 32 Zeitliche Änderung der Transmission der CH_3F -Ramanzelle (b) für CO_2 -Strahlung der Frequenz 1078.9 cm^{-1} . (a) zeigt die zeitliche Struktur des einfallenden bzw. des transmittierten Pulses.

Die Resonanzabsorption der CO₂-Laserstrahlung, die um so stärker wird, je näher die Frequenz der Strahlung dem Zentrum eines Rotations-Schwingungs-Übergangs $R(J)$ kommt, ist auch (zusammen mit einer Reabsorption der FIR-Photonen durch reine Rotationsübergänge im Schwingungsgrundzustand /15/), die Ursache für die starke Abnahme der FIR-Pulsenergie an den Rändern der einzelnen Abstimmintervalle innerhalb der Emissionscharakteristik, die in Widerspruch zu dem theoretisch dort erwarteten Anstieg der Kleinsignalverstärkung steht.

Mit Hilfe von Abb. 32 b läßt sich auch eine Abschätzung über den Wirkungsgrad der Umwandlung von CO₂-Laser-Photonen der Frequenz 1078.9 cm⁻¹ in FIR-Photonen der Frequenz 35.1 cm⁻¹ durchführen: Durch Vergleich der Transmission im Minimum (~ 20%) und der Transmission vor und nach dem Minimum (~ 35% bzw. ~ 60%) erhält man einen Wirkungsgrad zwischen 15% und 40% bezogen auf die Dauer des FIR-Pulses. Bezieht man den Wirkungsgrad auf den gesamten CO₂-Laserpuls, beträgt er zwischen 4% und 11%. Dies impliziert bei einer Pulsenergie der CO₂-Laserstrahlung von 250 mJ eine Pulsenergie für die FIR-Strahlung, die zwischen 300 μ J und 900 μ J liegt und somit in guter Übereinstimmung mit der Abschätzung steht, die mit Hilfe des Pyrodetektors P5-01 im vorhergehenden Abschnitt ermittelt wurde. (Die Skala für die FIR-Pulse in Abb. 31 wurde über einen Wirkungsgrad von 8% berechnet.)

Es sei zum Schluß noch bemerkt, daß stark strukturierte FIR-Pulse, wie sie in Abb. 31 gezeigt sind, natürlich nur auftraten, wenn die CO₂-Laserpulse entsprechend stark strukturiert waren. Eine mehr oder weniger glatte Form der CO₂-Laserpulse, wie sie gemäß den Ergebnissen in Abschnitt I.1.4.2 meist bei hohen Pulsenergien auftraten, übertrug sich auch auf die Form der FIR-Pulse. Beispiele dafür sind in den Abschnitten I.3.2 bzw. II.3.1 zu finden.

2.4 Übersicht über die Eigenschaften des CH₃F-Raman-Lasers

Die wesentlichen Eigenschaften des CH₃F-Raman-Lasers seien hier noch einmal kurz zusammengefaßt:

- Die abstimmbare Erzeugung von FIR-Strahlung ist im Frequenzbereich von 23 cm⁻¹ bis 45 cm⁻¹ in einzelnen Intervallen möglich, die eine Breite von 1.2 cm⁻¹ und untereinander einen Abstand von 0.4 cm⁻¹ besitzen. Innerhalb der Intervalle konnte eine Frequenzvariation in Schritten von 0.04 cm⁻¹ erfolgen.
- Die Bandbreite der emittierten FIR-Strahlung (FWHM) beträgt (4 ± 0.5) GHz. Durch Nachschalten eines Fabry-Perot-Interferometers konnte sie auf 400 MHz verringert werden.
- Pulsenergien von über 500 μ J sind erreichbar.
- Die Dauer der FIR-Pulse (FWHM) beträgt typisch 70 nsec.

Der hier vorgestellte Laser erscheint somit als ein geeignetes Instrument für die frequenz- und zeitabhängige Untersuchung nichtlinearer Phänomene im ferninfraroten Spektralbereich (z.B. erreicht man bei einer Fokussierung der FIR-Strahlung auf einen Durchmesser von 2 mm, siehe Abschnitt I.3.1, für Pulse der Energie $500 \mu\text{J}$ eine Spitzenintensität von $170 \frac{\text{kW}}{\text{cm}^2}$!). Er bietet zudem aber auch noch die Möglichkeit zur hochauflösenden Spektroskopie im herkömmlichen Sinn, und dies insbesondere an kleinen Proben (Durchmesser $< 2 \text{ mm}$) mit geringer Transmission ($< 1\%$), was mit handelsüblichen Fourierspektrometern nur sehr schwer möglich ist. Eine Demonstration dieser angesprochenen Möglichkeiten am Beispiel der nichtlinearen Absorption von FIR-Strahlung in Sauerstoff-dotiertem Silizium ist Gegenstand des II. Kapitels der vorliegenden Arbeit.

Verbesserungen der Eigenschaften des CH_3F -Raman-Lasers hängen im wesentlichen von Verbesserungen der Charakteristiken des Hochdruck- CO_2 -Lasers ab. Beispielsweise würde ein Betrieb des Hochdruck- CO_2 -Lasers auf nur einer longitudinalen Mode auch eine geringere Bandbreite der FIR-Strahlung bedingen. Dabei ist aber zu beachten, daß die Existenz verschiedener K -Zustände bei gegebener Quantenzahl J in symmetrischen Kreiselmolekülen stets zu einer Vielzahl eng benachbarter Emissionslinien führt, so daß die Reduzierungsmöglichkeiten der Bandbreite des FIR-Lasers begrenzt sind /63/. Zur Erweiterung der erzeugbaren FIR-Frequenzen wären andere molekulare Gase, wie das bereits besprochene $^{13}\text{CH}_3\text{F}$, und die vom konventionellen Laserbetrieb her bekannten Gase CD_3F , CH_3Cl , CH_3I oder NH_3 zu untersuchen.

1.3. Experimentelle Anordnung und Techniken für die FIR-Transmissionspektroskopie mit Hilfe des CH_3F -Raman-Lasers

Mit Hilfe des abstimmbaren CH_3F -Raman-Lasers wurde eine Anordnung zur Untersuchung der FIR-Transmission von Festkörpern bei tiefen Temperaturen aufgebaut. Dieser Abschnitt schildert einige wesentliche Aspekte des optischen Aufbaus, des Meßdatenerfassungssystems sowie der verwendeten experimentellen Techniken.

3.1 Optischer Aufbau

Das Ziel der optischen Anordnung, die in Abb. 33 gezeigt ist, war, eine maximale Intensität der FIR-Strahlung am Ort der Probe, die sich im Zentrum eines Heliumbad-Kryostaten befand, zu erreichen. Zu beachtende Randbedingungen waren dabei zum einen die endliche Divergenz der FIR-Strahlung, die an der Austrittsöffnung des Quarzglaswellenleiterrohres gebeugt wird, und zum anderen der Öffnungswinkel des Kryostaten, der durch den Durchmesser der Kryostatenfenster und deren Abstand von der Probe gegeben war. Diese Randbedingungen erlaubten eine Abbildung der Austrittsöffnung des Quarzglaswellenleiters auf die Probe mit einer maximalen Verkleinerung von 1:4. Die Abbildung erfolgte in zwei Schritten: Zunächst wurde mittels des konkaven Spiegels K1 ein bereits 1:4 verkleinertes Zwischenbild der Austrittsöffnung am Orte P1 erzeugt, das dann mit Hilfe des konkaven Spiegels K2 im Verhältnis 1:1 auf die Probe abgebildet wurde. Diese doppelte Abbildung war ebenso wie die zusätzliche Umlenkung durch die ebenen Spiegel E1 und E2 nötig, um die Winkel zwischen einfallenden und reflektierten Strahlen an den beiden Hohlspiegeln so gering wie möglich ($\leq 10^\circ$) zu halten. Ansonsten hätte die jeweilige Ausbildung eines Komas, d.h. einer Verzerrung des Bildes infolge des außeraxialen Einfalls eines Teils der Strahlung auf die Hohlspiegel, eine zu starke Reduktion der Intensität der FIR-Strahlung am Orte der Probe zur Folge gehabt. (Nichtsphärische Spiegel oder FIR-transparente Linsen, die eine direkte verkleinerte Abbildung der Austrittsöffnung auf die Probe erlaubt hätten, standen nicht zur Verfügung.) Da die Schwankungen der Pulsenergie der FIR-Strahlung, insbesondere an den Rändern der einzelnen Abstimmintervalle, relativ groß waren, war es unerlässlich, einen Teil der einfallenden FIR-Strahlung zur Kontrolle der Pulsenergie über einen Strahlteiler T1 auf einen Detektor D_{Ref} auszukoppeln. Als Strahlteiler wurde eine dünne gespannte Polyäthylenfolie (Dicke $25\ \mu\text{m}$) verwendet. Die Anordnung der Strahlteilerfolie und des Referenzdetektors D_{Ref} war so gewählt, daß auch am Orte des Detektors ein scharfes, im Verhältnis 1:4 verkleinertes Bild der Austrittsöffnung des Wellenleiterrohres vorlag. Dasselbe galt für den Detektor D_{T} , der zur Beobachtung der transmittierten FIR-Strahlung diente; denn der Querschnitt der durch die Probe transmittierten Strahlung wurde mit Hilfe des

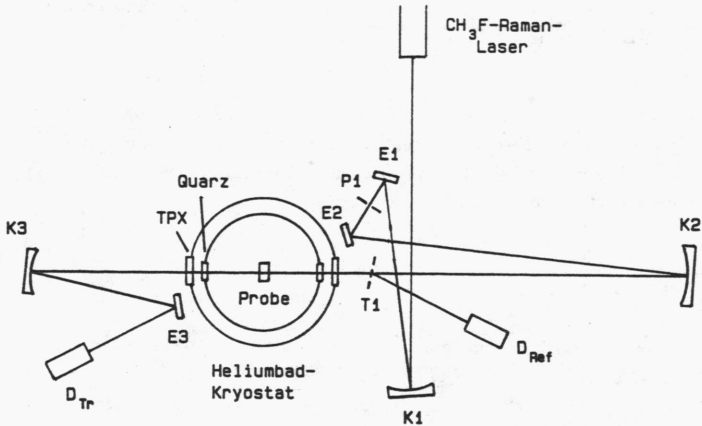


Abb. 33 Skizze des optischen Aufbaus für die Transmissionsspektroskopie mit Hilfe des CH_3F -Raman-Lasers (nicht maßstabsgerecht; Erläuterung siehe Text).

Hohlspiegels K3 wiederum im Verhältnis 1:1 auf den Detektor D_{Tr} abgebildet. Die Durchmesser sämtlicher Spiegel waren selbstverständlich dem Durchmesser der FIR-Strahlung am Orte der Spiegel angepaßt.

Die Querschnittsverteilung der Energiedichte der FIR-Strahlung am Orte der Probe (und nach der obigen Diskussion auch am Orte der Detektoren) ließ sich bei der gezeigten Anordnung annähernd durch eine Gaußsche Verteilung mit einer Halbwertsbreite (FWHM) von $\Delta_p \simeq 1.7 \text{ mm}$ beschreiben. Zur quantitativen Angabe der Energiedichteverteilung ist die Kenntnis der Verluste der FIR-Strahlung auf dem Weg vom Austrittsfenster der Ramanzelle - dort habe sie die Pulsenergie E_0 - bis zur Probe nötig. Diese wurden durch die Reflexionsverluste am Innenfenster des Kryostaten (Quarz, 3 mm dick, Transmission für Strahlung von 29.2 cm^{-1} : 70%), am Außenfenster des Kryostaten (TPX, 3 mm dick, Transmission 85%) und am Strahlteiler T1 (Polyäthylenfolie, $\sim 25 \mu\text{m}$ dick, Transmission 90%) verursacht (Absorptionsverluste in Luft können bei einer Frequenz von 29.2 cm^{-1} vernachlässigt werden). Für die Energiedichte \tilde{E} im Zentrum des Laserstrahls am Probenort gilt dann die Bedingung:

$$0.5 E_0 = \int_0^\infty \int_0^\infty \tilde{E} \exp\left(\frac{-4 \ln 2}{\Delta_p^2}(x^2 + y^2)\right) dx dy \quad (37a)$$

oder

$$\tilde{E} = \frac{2 \ln 2}{\pi} \frac{E_0}{\Delta_p^2} = \frac{2}{\pi} \frac{\Delta^2}{D_p^2 \Delta_p^2} \cdot \frac{U}{R_0} \cdot \gamma \quad (37b)$$

Die letzte Identität folgt unter Verwendung von Gl.(36). Dort sind auch die Größen Δ , D_p , U , R_0 und γ definiert. Einsetzen der Größen liefert $\tilde{E} = 0.08 \cdot \frac{\text{J}}{\text{cm}^2} \cdot \gamma U$. U ist die Spitzenspannung des Pyrodetektors P5-01 in der Energiemeßmode und γ der in Abschnitt I.2.2 besprochene frequenzabhängige Korrekturfaktor. Gl.(37) war die Grundlage für die Bestimmung der Energiedichteverteilung der FIR-Strahlung am Probenort bei den in Kapitel II besprochenen Experimenten.

Zur Messung der Pulsenergie von einfallender und transmittierter Strahlung wurden zwei Golayzellen verwendet. Da die als Referenzdetektor D_{ref} eingesetzte Golayzelle ein Quarzfenster besaß, konnte auf die in Abschnitt I.2.2 erwähnte Quarzscheibe, die die restliche das TPX-Austrittsfenster der Ramanzelle durchdringende CO_2 -Laserstrahlung abblocken sollte, entfernt werden. (Auch das Innenfenster des Kryostaten war aus Quarz!) Die immerhin 30% betragenden Reflexionsverluste der FIR-Strahlung an der Scheibe konnten somit vermieden werden. Der Durchmesser der Eintrittsfenster der beiden Golayzellen betrug jeweils 3 mm, so daß bei einer optimalen Justierung nahezu die gesamte auftreffende Strahlung ($\Delta_p = 1.7$ mm) von den Golayzellen erfaßt wurde.

Um die zeitliche Pulsstruktur von einfallender und transmittierter Strahlung zu untersuchen, wurden die Golayzellen durch zwei Schottkydioden (SD017 der Fa. Farran Technology) ersetzt. Da der Antennenquerschnitt dieser Detektoren, der etwa als λ_L^2 angenommen werden kann (λ_L = Wellenlänge der FIR-Strahlung) klein gegen die Halbwertsbreite Δ_p der Intensitätsquerschnittsverteilung war, konnte mit ihnen nur die Intensität an einem Ort der Verteilung - gewählt wurde natürlich das Zentrum - beobachtet werden. Um die Detektoren nicht zu überlasten, wurden vor ihnen Abschwächer für die FIR-Strahlung angebracht.

Ein Problem, das vor allem bei der Untersuchung von Proben mit einem Durchmesser unter 3 mm auftrat, war eine starke Streuung des Verhältnisses der Signale von Transmissions- und Referenzdetektor von Puls zu Puls, die auch bei konstant gehaltener Frequenz und Pulsenergie der FIR-Strahlung bis zu 20% betragen konnte. Ursache dieses Verhaltens war eine von Puls zu Puls verschiedene Modenausbreitung im Quarzglaswellenleiter des FIR-Lasers wegen der von Puls zu Puls verschiedenen Verteilung longitudinaler Moden innerhalb der Bandbreite des CO_2 -Lasers. Die Folge waren leicht unterschiedliche Ausbreitungsrichtungen der FIR-Strahlung und damit räumliche Schwankungen des Bildes der Austrittsöffnung des Quarzglaswellenleiters am Orte der Probe bzw. der Detektoren. Bei Probendurchmessern ≤ 3 mm trat dann trotz konstanter Einfallspulsenergie eine Streuung in der Energie des transmittierten Pulses auf. Denn auch bei einer Halbwertsbreite $\Delta_p = 1.7$ mm beträgt die 10% Breite der Querschnittsverteilung der Energiedichte immerhin schon 3 mm! Natürlich trägt auch der endliche Durchmesser der Eintrittsfenster der Golayzellen (3 mm!)

zu den beobachteten Schwankungen bei. Abhilfe war möglich durch Einfügen einer Blende (Öffnungsdurchmesser 1 mm) am Orte des ersten Zwischenbildes P1 in Abb. 33. Die Ausbreitung der FIR-Strahlung nach der Blende wurde dann durch die Beugung der Strahlung an der Blende bestimmt und war damit unabhängig von den Fluktuationen der Moden der FIR-Strahlung. Die Schwankungen des Verhältnisses der Signale der Detektoren D_T und D_{ref} waren dann typischerweise $< 3\%$. Für eine präzise Messung des Transmissionsspektrums einer Probe war das Einfügen dieser Blende unerlässlich. Infolge der starken Beugung an der Blende war jedoch der Durchmesser der Spiegel nicht mehr groß genug, um die gesamte Strahlung aufzufangen. Die Energiedichte der Strahlung am Ort der Probe war geringer als nach Gl.(37) folgen würde. Bei der Untersuchung der Abhängigkeit der Transmission einer Probe von der Pulsenergie, wie sie in Kapitel II dieser Arbeit vorgestellt wird, wurde die Blende daher entfernt.

3.2 Meßwerterfassungssystem

Abb. 34 zeigt eine schematische Skizze der Meßelektronik zur Bestimmung der Energietransmission einer Probe, d.h. des Verhältnisses der Energie des durch die Probe transmittierten und des auf die Probe auftreffenden FIR-Pulses. Die Prozeßsteuerung, Meßdatenerfassung und -Auswertung wurde durch einen Z80-ECB-Bus-Rechner übernommen. Der Rechner war mit 2 Floppy-Disk-Laufwerken, einem Plotter (HP 7220C), einer Graphikkarte mit Bildschirm (512x512 Pixel) und diversen Prozeßinterfaces (serielle und parallele Schnittstellen, ADC's, DAC's) ausgerüstet.

Die Reaktion einer Golayzelle auf einen einfallenden FIR-Puls ($\text{FWHM} \approx 70 \text{ nsec}$) bestand in einem etwa 100 msec langen Puls negativer Polarität. Die Pulshöhe war ein Maß für die Energie des auftreffenden FIR-Pulses. Ein linearer Zusammenhang bestand in einem Bereich von 50 mV bis 7 V; der Dynamikbereich einer Golayzelle betrug also über 2 Dekaden. Die untere Grenze war durch das Rauschen der Golayzelle, das im wesentlichen durch Fluktuationen der Umgebungstemperatur verursacht wurde, bedingt. Die Aufzeichnung der von der Golayzellen gelieferten Signale erfolgte mit Hilfe eines 2-Kanal-Transientenrekorders (Data-lab DL 922). Dieser besitzt eine Amplitudenauflösung von 8 bit und eine Speichertiefe von 2 kByte pro Kanal; die Abtastrate kann von 250 Hz bis 10 MHz, die Eingangsempfindlichkeit von 100 mV bis 50 V variiert werden. Die Eingangsbandbreite beträgt 6 MHz.

Der Vorgang einer Transmissionsmessung war wie folgt: Über ein Schrittmotortreiberinterface wurde vom Rechner die gewünschte Emissionsfrequenz des Hochdruck- CO_2 -Lasers und damit auch des CH_3F -Raman-Lasers eingestellt (oder alternativ der Gitterabstand des in Abschnitt I.2.2 besprochenen, dem CH_3F -Raman-Laser nachgeschalteten Fabry-Perot-Interferometers verändert). Daraufhin wurde der Transientenrekorder, der in der Pretrigger-

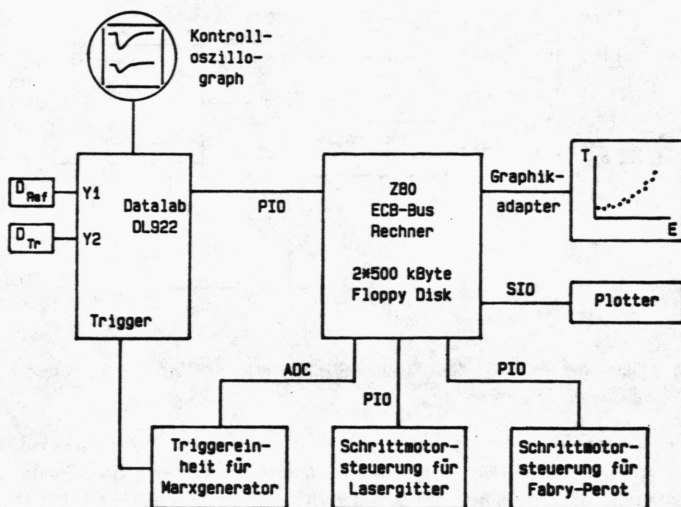


Abb. 34 Prozesssteuerung, Datenerfassung und -Auswertung für Messung der Energie-
transmission.

mode betrieben wurde, um die Nulllinie vor dem Eintreffen der Golayzellensignale erfassen zu können, vom Rechner armiert und die Entladung des Hochdruck-CO₂-Lasers gezündet. Die Triggerpulse für den Transientenrekorder wurden von einem Synchronisationsausgang der Marxgeneratortriggereinheit geliefert. Nach Ende des Digitalisierungsprozesses wurde der Speicherinhalt des Transientenrekorders, der zur visuellen Kontrolle der Golayzellensignale auch auf einem Oszillographen beobachtbar war, über eine parallele Schnittstelle vom Rechner übernommen. Die Höhe der von den Golayzellen gelieferten Signale wurde ermittelt und die Transmission berechnet. (Der genaue Rechengang unter Berücksichtigung der notwendigen Eichprozeduren wird in Abschnitt I.3.3 beschrieben.)

Eine Mittelung über mehrere Pulse war möglich; Mittelwert und Standardabweichung wurden automatisch berechnet. Signale, die aufgrund von Schwankungen der FIR-Pulsenergie die Eingangsverstärker des Transientenrekorders sättigten, wurden vom Rechner ignoriert. Eine Darstellung der Resultate, d.h. der Transmission der Probe als Funktion der Frequenz oder der Pulsenergie der FIR-Strahlung, war online auf dem Graphikbildschirm oder nach Beendigung einer Meßserie am Plotter möglich.

Um die zeitliche Änderung der Transmission der Probe während des FIR-Laserpulses erfassen zu können, wurden die Golayzellen durch die erwähnten Schottkydioden ersetzt.

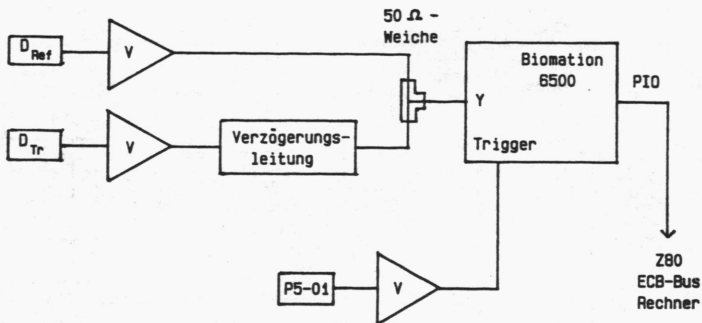


Abb. 35 Simultane Erfassung der zeitlichen Struktur des einfallenden und des transmittierten FIR-Pulses.

Statt des 2-Kanal-Transientenrekorders, der für die Aufzeichnung der Pulsstrukturen der einfallenden und transmittierten Strahlung zu langsam war, wurde ein schneller 1-Kanal-Transientenrekorder verwendet (Biomation 6500, Amplitudenauflösung 6 bit, Abtastrate 500 MHz, Eingangsbandbreite 125 MHz, Eingangsimpedanz 50 Ω). Um trotzdem die Signale beider Schottkydioden während ein und desselben Laserpulses aufzeichnen zu können, wurde zu dem in Abb. 35 gezeigten Trick gegriffen: Die Signale der Schottkydioden wurden verstärkt (Verstärker: Ortec 535: Bandbreite DC-120 MHz, Ausgangsimpedanz 50 Ω ; Plessey SL560: Bandbreite 100 MHz, Ausgangsimpedanz 50 Ω) und über eine 50 Ω -Weiche an den Eingang des Transientenrekorders angelegt. Das Signal der zur Detektion der transmittierten Strahlung verwendeten Schottkydiode wurde hierbei mit Kabeln (Wellenwiderstand 50 Ω) um etwa 250 nsec verzögert. Die genaue Verzögerungszeit wurde mit einem schnellen Pulsgenerator (Philips PM 5786) ausgemessen. Ein Problem dieser Anordnung war, daß die Verzögerungsleitung zu einem Verschleifen der Signale der Schottkydiode D_{Tr} führte. Daß das Meßverfahren dennoch akzeptable Resultate lieferte, zeigt die zeitlich konstante Transmission in Abb. 36 bei einer Messung ohne Probe (die FIR-Frequenz war 29.22 cm^{-1} ; diese Messung war die Referenzmessung zu der in Abschnitt II.3.1 gezeigten Messung mit Probe).

Die eigentliche Durchführung der Messung der zeitabhängigen Transmission erfolgte ganz analog der oben beschriebenen Durchführung einer Energietransmissionsmessung. Der einzige Unterschied war, daß eine Triggierung des schnellen Transientenrekorders über den Synchronisationsausgang der Marxgeneratortriggereinheit infolge des großen zeitlichen Jitters zwischen dem Synchronisationspuls und der Emission des FIR-Pulses, der bis zu 1 μsec betragen konnte, nicht mehr möglich war. Die Erzeugung des Triggersignals erfolgte daher durch Teilauskopplung der CO_2 -Laserstrahlung auf einen pyroelektrischen Detektor (P5-01, Anstiegszeit 0.5 nsec), dessen Signale verstärkt wurden und auf den Triggereingang des

Transientenrekorders gelegt wurden.

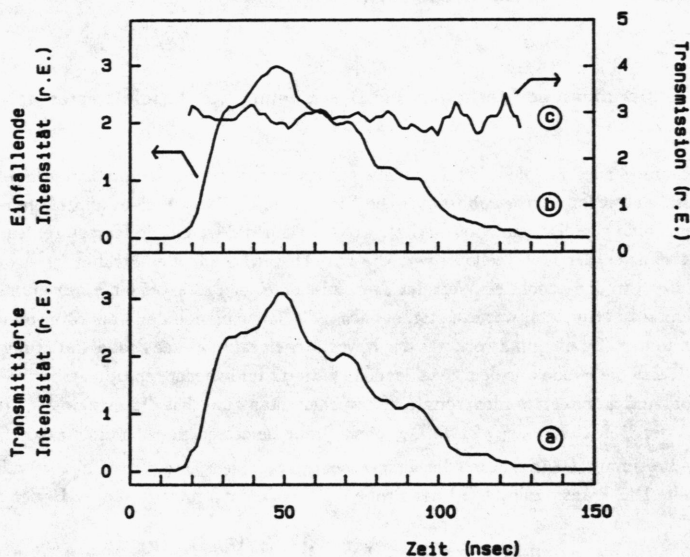


Abb. 36 Signale der Schottkydioden ((a) Referenzdiode, (b) Transmissionsdiode) und daraus bestimmte Transmission (c) für den Fall, daß sich keine Probe im Strahlengang befand ($\nu_{FIR} = 29.22 \text{ cm}^{-1}$).

Die in Abschnitt I.2.3.2 besprochenen Experimente zur Bestimmung des Wirkungsgrades des CH_3F -Raman-Lasers wurden mit entsprechend anderen Detektoren ebenfalls auf die hier beschriebene Methode durchgeführt.

Es sei noch erwähnt, daß die Programme für den Rechner in der Sprache MILAN, die eine Modifikation der Sprache PASCAL darstellt und die vom Rechenzentrum der Universität Regensburg speziell für Echtzeitanwendungen im Labor entwickelt wurde, geschrieben wurden. Die zeitkritischen und direkt auf die Hardware bezogenen Aufgaben, wie z.B. das Auslesen der Daten aus den Transientenrekordern, waren direkt in der Form von Assemblerunterprogrammen implementiert. Außerdem sei vorweggenommen, daß alle in Kapitel II dieser Arbeit vorgestellten Modellrechnungen für die nichtlineare Absorption von FIR-Strahlung in Sauerstoff-dotiertem Silizium auf einem IBM-XT-kompatiblen Rechner (Olivetti M24) in der Programmiersprache TURBO_C durchgeführt wurden. Rechenintensive Routinen

wurden dabei durch Direktprogrammierung des Arithmetik-Coprozessors 8087 in Assembler bezüglich ihrer Geschwindigkeit optimiert.

3.3 Experimentelle Methoden zur Bestimmung des Absolutwertes der Transmission

Ein grundlegendes Problem bei Transmissionsbestimmungen im ferninfraroten Spektralbereich ist, daß jedes durchstrahlte optische Element infolge von Vielstrahlinterferenzen eine frequenzabhängige Transmission besitzt. Solche Elemente waren die Kryostatfenster, die Eintrittsfenster der Golayzellen sowie die zur Abschwächung der Strahlung verwendeten Elemente. Um den absoluten Wert der Transmission einer Probe bei einer bestimmten FIR-Frequenz ν zu ermitteln, war es nötig, zunächst die Transmission der Apparatur ohne Probe zu bestimmen. Zu diesem Zweck war der Kryostateneinsatz mit der Probenhalterung so konstruiert, daß die Probe aus dem Strahlengang entfernt und wieder reproduzierbar an die alte Position zurückbewegt werden konnte. Es wurden dann zunächst die Signale $S_{Ref}^0(\nu)$ bzw. $S_{Tr}^0(\nu)$ der Detektoren D_{Ref} bzw. D_{Tr} ohne Probe gemessen und der Eichfaktor $\mathcal{E}f(\nu) = \frac{S_{Ref}^0(\nu)}{S_{Tr}^0(\nu)}$ bestimmt. Dann wurden die entsprechenden Signale $S_{Ref}(\nu)$ bzw. $S_{Tr}(\nu)$ mit Probe ermittelt. Die Transmission $T(\nu)$ der Probe ergibt sich dann nach folgender Beziehung:

$$T(\nu) = \mathcal{E}f(\nu) \cdot \frac{S_{Tr}(\nu)}{S_{Ref}(\nu)} = \frac{S_{Ref}^0(\nu)}{S_{Tr}^0(\nu)} \frac{S_{Tr}(\nu)}{S_{Ref}(\nu)} \quad (38)$$

Bei der Aufnahme des Transmissionsspektrums einer Probe wurde diese Prozedur für jede Frequenz des CH_3F -Raman-Lasers durchgeführt. Üblicherweise wurde bei Energietransmissionsmessungen sowohl bei der Bestimmung des Eichfaktors als auch der Transmission über mehrere Pulse gemittelt. Bei Beobachtung der zeitlichen Änderung der Transmission während des FIR-Pulses war eine Mittelung wegen der sich von Puls zu Puls ändernden Struktur nicht sinnvoll.

Einige Bemerkungen zur Meßmethode, die bei Untersuchungen der nichtlinearen Energietransmission einer Probe, wie sie in Kapitel II dieser Arbeit vorgestellt werden, verwendet wurde, seien noch angeführt. Hier wurde der Eichfaktor nur zu Beginn einer Meßserie ermittelt. Die Transmission der Probe bei verschiedenen Pulsenergien wurde dann stets nach Gl.(38) berechnet. Änderungen im optischen Strahlengang, wie z.B. eine Variation der Abschwächer vor den Detektoren, durften während einer Meßserie nicht mehr erfolgen. Der maximale Änderungsbereich der Pulsenergie war damit durch den Dynamikbereich der Golayzellen (etwa 2 Dekaden) gegeben. Bei größeren Änderungen der Pulsenergie mußte eine neue Meßserie mit neuen Abschwächern vor den Detektoren und damit mit einem neuen Eichfaktor durchgeführt werden.

Es erwies sich auch als ungünstig, eine Variation der FIR-Pulsenergie durch Abschwächer nach dem Austrittsfenster des CH_3F -Raman-Lasers durchzuführen. Denn die verschiedenen Moden, die in der Bandbreite der emittierten Strahlung enthalten waren, wurden von den Abschwächern wiederum wegen Vielstrahlinterferenzen innerhalb und zwischen den Abschwächern unterschiedlich stark gedämpft. Die spektrale Energieverteilung der Laserstrahlung nach den Abschwächern war damit abhängig von ihrer Art und Anzahl. Da aber der Eichfaktor nur für eine solche spektrale Energieverteilung aufgenommen wurde, ergaben sich somit starke Streuungen im Absolutwert der Transmission. Als beste Methode für die Messung der Abhängigkeit der Proben-Transmission von der Pulsenergie erwies sich folgendes Verfahren:

Der CH_3F -Raman-Laser wurde auf maximale Pulsenergie eingestellt und damit der Eichfaktor bestimmt. Eine Reduktion der Pulsenergie ließ sich auf einfache Weise durch langsame Erhöhung des CH_3F -Gasdruckes in der Ramanzelle erreichen. Durchgeführte Tests ergaben, daß der Eichfaktor dann über den gesamten Dynamikbereich der Golayzellen tatsächlich unabhängig von der Energie der FIR-Strahlung blieb (dies war insbesondere auch ein Beweis für die Linearität der Golayzellen innerhalb ihres Dynamikbereichs). Während der Druckerhöhung wurde dann kontinuierlich (d.h. mit der üblichen Wiederholrate für die Zündung der Entladungen des Hochdruck- CO_2 -Lasers von 0.5 Hz) die Transmission der Probe gemessen und vom Rechner als Funktion der auf die Probe fallenden Pulsenergie aufgetragen (eine Mittelung erfolgte hierbei nicht). Ein relatives Maß für die Pulsenergie war dabei das Signal der Referenzgolayzelle; eine absolute Energieeichung erfolgte durch Vergleich der Signale der Referenzgolayzelle mit denen des Pyrodetektors P5-01 in der Energiemeßmode unter Zuhilfenahme der Gln.(36) und (37).

II. Nichtlineare Absorption von Ferninfrarotstrahlung durch eine resonante Bandmode in Sauerstoff-dotierten Siliziumkristallen

Die Eignung des in den vorhergehenden Abschnitten beschriebenen Lasersystems für die Untersuchung nichtlinearer Phänomene im fernen Infrarot soll in diesem Kapitel anhand der erstmals beobachteten nichtlinearen Absorption von FIR-Strahlung in Sauerstoff-dotiertem Silizium (in Zukunft mit Si:O bezeichnet) demonstriert werden. Aus Absorptionsmessungen mittels Fourierspektrometern ist bekannt, daß Sauerstoff-dotiertes Silizium eine schmalbandige Resonanz bei 29.2 cm^{-1} im Frequenzbereich der akustischen Phononen besitzt, verursacht durch die Anregung einer sog. resonanten Bandmode (siehe die folgenden Abschnitte). Da diese Resonanzfrequenz in unmittelbarer Nachbarschaft eines Emissionsmaximums des CH_3F -Raman-Lasers bei 29.35 cm^{-1} liegt, ist eine Untersuchung der resonanten Bandmode bei hoher, nahezu monochromatischer Anregung durch FIR-Strahlung möglich. Untersuchungsgegenstand in dieser Arbeit war die Transmission eines Si:O-Kristalls in der Umgebung der Resonanzfrequenz bei tiefen Temperaturen, die eine starke Zunahme mit der Pulsenergie des FIR-Lasers aufwies. Ziel war es, aus einer Analyse der nichtlinearen Transmission Rückschlüsse auf mögliche Dämpfungsmechanismen für die resonante Bandmode und auf zugehörige Relaxationszeiten zu ziehen.

Bevor nun in den folgenden Abschnitten die Meßergebnisse vorgestellt und interpretiert werden, sollen zunächst das System Si:O und seine Absorptionseigenschaften im fernen Infrarot behandelt werden.

II.1 Interstitieller Sauerstoff in Siliziumkristallen

1.1 Bedeutung von Sauerstoff als Verunreinigung in Silizium bei der Produktion von Halbleiterbauelementen

Hier sei kurz erläutert, warum Sauerstoff bzw. durch Sauerstoff verursachte Defekte in Silizium seit langem Gegenstände besonderen Interesses bei der Herstellung von höchstintegrierten (VLSI) Schaltkreisen sind (detaillierte Informationen finden sich in [90,91]).

80% des in der Halbleiterindustrie verwendeten Siliziums werden nach dem Czochralski-Verfahren, d.h. aus der Schmelze gezogen und werden damit durch Reduktionsprozesse

an den dabei verwendeten Quarztiegeln mit Sauerstoff bis zu einer Konzentration von $2 \cdot 10^{18} \text{ cm}^{-3}$ verunreinigt /92/. (Zonengezogenes Silizium enthält dagegen nur eine Sauerstoffkonzentration von weniger als 10^{16} cm^{-3} .) Sauerstoffkontamination kann aber auch erfolgen, wenn teilweise mit Oxidschichten bedeckte Siliziumwafer (beispielsweise bei der Herstellung von MOS-Schaltkreisen) hohen Temperaturen ausgesetzt werden.

Wie im folgenden Abschnitt diskutiert, liegt der Sauerstoff in Silizium im wesentlichen in elektrisch inaktiver interstitieller (d.h. auf Zwischengitterplätzen eingelagerter) Form vor und zeigt keinen Einfluß auf die Funktionsweise von Bauelementen. Bei hohen Temperaturen, wie sie während des Herstellungsprozesses von VLSI-Schaltkreisen auftreten, wird der Sauerstoff jedoch beweglich und es bilden sich zusätzliche Defekte, die sich positiv wie negativ auf die Funktionsweise der Schaltkreise auswirken können. Als Beispiele seien SiO_2 -Einlagerungen (Präzipitate), die bis zu $40 \mu\text{m}$ Durchmesser erreichen können, sowie sog. thermische Donatoren /93/ genannt: SiO_2 -Einlagerungen verursachen Versetzungen und Stapelfehler. An der Oberfläche bewirken diese beispielsweise bei der Herstellung von Bipolartransistoren, wo Basisdicken von $\sim 1 \mu\text{m}$ gefordert werden, eine Verschlechterung der Transistoreigenschaften (z.B. Reduktion der Stromverstärkung infolge Störstellenanlagerung an den Versetzungen). Im Substrat sorgen sie jedoch wegen ihrer Getterwirkung für Verunreinigungen für defektfreie Oberflächen /93/. Thermische Donatoren sind Silizium-Sauerstoff-Komplexe der Form Si_mO_n , die bei hohen Temperaturen entstehen und auch nach Abkühlen erhalten bleiben. Diese Komplexe bilden Doppeldonatoren /94/ und ändern somit die Ladungsträgerkonzentration in den Schaltelementen während der Produktion.

Die genannten Defekte sind jedoch nicht Gegenstand der Untersuchungen dieser Arbeit. Hier werden vielmehr elementare Eigenschaften der Ausgangsform der Defekte, nämlich der interstitiellen Form des Sauerstoffs in Siliziumkristallen, behandelt.

1.2 Einbaulage von Sauerstoff in Siliziumkristallen

Daß Sauerstoff in Siliziumkristallen nicht substitutionell, sondern interstitiell eingebaut ist, zeigten Bond und Kaiser /95/ anhand von Röntgenbeugungsmessungen (mit Sauerstoff dotierte Siliziumkristalle zeigen eine größere Gitterkonstante als undotierte) zusammen mit Dichteuntersuchungen.

Die heute allgemein anerkannte Modellvorstellung für den Einbau des interstitiellen Sauerstoffs in Siliziumkristallen ist folgende /96/: Zwei nächstbenachbarte Si-Atome in [111]-Richtung geben ihre kovalente Bindung miteinander auf und gehen statt dessen eine kovalente Bindung mit einem O-Atom ein, das in der Mitte zwischen beiden Si-Atomen, aber etwas außerhalb ihrer Verbindungslinie lokalisiert ist. Eine Illustration mit den charakteristischen Größen zeigt Abb. 37.

$$\alpha = 162^\circ$$

$$d = 2.35 \text{ \AA}$$

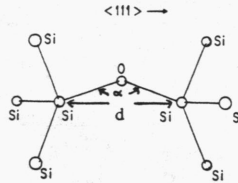


Abb. 37 Einbaulage des interstitiellen Sauerstoffs in Silizium.

1.3 Infrarotabsorptionseigenschaften von Sauerstoff-dotiertem Silizium

Die im vorigen Abschnitt beschriebene dreieckige Anordnung Si-O-Si läßt sich in erster Näherung (bei Vernachlässigung jeglicher Wechselwirkung mit dem restlichen Siliziumgitter) als ein in den Siliziumkristall eingebettetes Si_2O -Molekül betrachten, vergleichbar dem Si_2O -Komplex in gasförmigem Disiloxan $\text{H}_3\text{SiOSiH}_3$. Ein solches nichtlineares dreiatomiges Molekül besitzt bekanntlich 3 Eigenschwingungen /97/ (Abb. 38): die symmetrische Streckschwingung (ν_1), die symmetrische Biegeschwingung (ν_2) und die antisymmetrische Streckschwingung (ν_3), die allesamt ein nicht verschwindendes elektrisches Dipolmoment besitzen und daher infrarotaktiv sind.

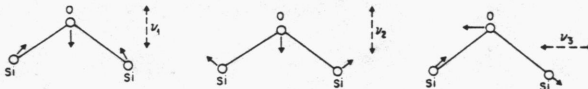


Abb. 38 Normalschwingungen eines freien Si_2O -Moleküls. (Die gestrichelten Pfeile geben die Richtung des elektrischen Dipolmoments an.)

Tatsächlich werden in Sauerstoff-dotierten Siliziumkristallen 3 Absorptionsbande im mittleren Infrarot beobachtet, nämlich bei $8 \mu\text{m}$ (1203 cm^{-1}), $9 \mu\text{m}$ (1136 cm^{-1}) und bei $19 \mu\text{m}$ (517 cm^{-1}) /97/. Die Zuordnung der $9 \mu\text{m}$ -Absorption an die ν_3 -Schwingung ist konsistent mit Absorptionsmessungen an gasförmigem Disiloxan. Eine Zuordnung der Absorptionsbande bei $8 \mu\text{m}$ bzw. $19 \mu\text{m}$ an die symmetrische Streckschwingung ν_1 bzw. die symmetrische Biegeschwingung ν_2 , die die geringste Energie haben muß, ist jedoch nicht möglich. Denn

bei Siloxanen ist die Energie der ν_1 -Schwingung stets viel geringer als die der ν_3 -Schwingung /98/. Außerdem würde die Frequenz der ν_2 -Biegeschwingung bereits unterhalb der Ramanfrequenz (519 cm^{-1} /99/) des reinen Siliziumgitters im Bereich der optischen Phononen liegen, so daß die Vorstellung eines freien Si_2O -Moleküls nicht mehr gültig ist. Auch die bei tiefen Temperaturen beobachtete komplizierte Feinstruktur der $9\text{ }\mu\text{m}$ -Absorptionsbande ist mit dieser Vorstellung nicht erklärbar /96/.

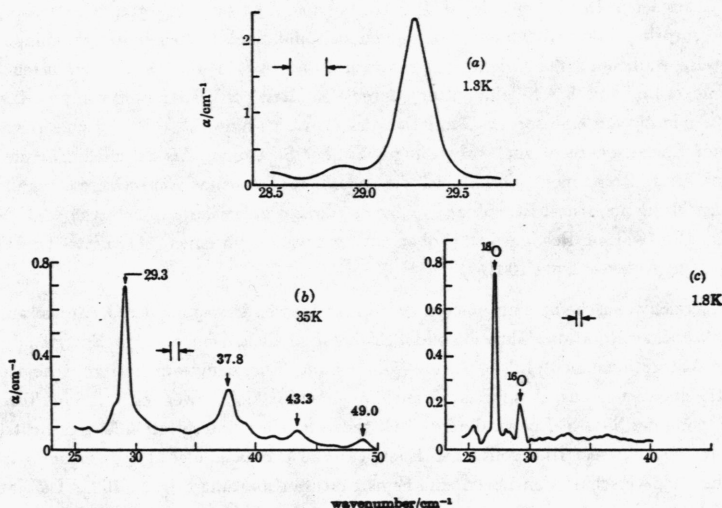


Abb. 39

- Absorptionslinie von ^{16}O (Konzentration $7 \cdot 10^{17}\text{ cm}^{-3}$) in Silizium bei $T = 1.8\text{ K}$.
- Absorptionsspektrum von ^{16}O ($7 \cdot 10^{17}\text{ cm}^{-3}$) in Silizium bei $T = 35\text{ K}$.
- Absorptionslinien von ^{18}O ($2 \cdot 10^{17}\text{ cm}^{-3}$) und ^{16}O ($5 \cdot 10^{16}\text{ cm}^{-3}$) in Silizium bei $T = 1.8\text{ K}$ (nach /100/).

Ein modifiziertes Modell wurde von Bosomworth et al. /100/ aufgrund von Absorptionsmessungen im Bereich des fernen Infrarot vorgeschlagen. Denn sie fanden bei einer Temperatur von 1.8 K eine Absorptionslinie bei 29.3 cm^{-1} , zu der bei höheren Temperaturen ($> 35\text{ K}$) noch Linien bei 37.8 cm^{-1} , 43.3 cm^{-1} und 49 cm^{-1} hinzukamen, was auf Übergänge von angeregten Zuständen in noch höhergelegene Zustände hindeutet (Abb. 39 a,b). Experimente an zusätzlich mit dem Sauerstoffisotop ^{18}O dotierten Si-Kristallen ergaben neben der Absorption bei 29.3 cm^{-1} eine weitere bei 27.2 cm^{-1} , was eindeutig auf Sauerstoff als Ursache

für die Absorptionslinien hinweist. Die Existenz dieser beiden Linien wurde auch durch Phononenspektroskopie mittels supraleitenden Tunnelkontakten bestätigt /101/. Die Breite der Absorptionslinie bei 29.3 cm^{-1} in Abb. 39 a ist nach Bosomworth durch die Auflösung des verwendeten Spektrometers ($= 0.2 \text{ cm}^{-1}$) gegeben.

Bosomworth et al. betrachteten in ihrem Modell nur eine Bewegung des interstitiellen Sauerstoffatoms in einem in erster Näherung als starr angenommenen Siliziumgitter. Das O-Atom besitzt somit 3 Freiheitsgrade für eine Schwingung gegenüber dem Siliziumgitter. Experimente, bei denen Druck längs der [111]-Richtung ausgeübt wurde, zeigten eine Aufspaltung der Absorptionslinie bei 29.3 cm^{-1} in eine einfach und eine dreifach entartete Komponente (teilweise Aufhebung der Orientierungsentartung des in [111]-Richtung orientierten Si_2O -Komplexes im kubischen Siliziumgitter), deren Verhalten bei Verwendung von polarisiertem Licht eindeutig auf eine sog. σ -Oszillation des O-Atoms hinweist, d.h. auf eine Bewegung in einer Ebene senkrecht zur Verbindungsachse der Si-Atome. Als Beispiel: Für die dreifach entartete Komponente verhält sich die absorbierte Intensität von Licht, das parallel zur Druckrichtung polarisiert ist, zu der von dazu senkrecht polarisiertem Licht wie 8:5 für einen σ -Oszillator (wie von Bosomworth beobachtet) und wie 1:4 für einen π -Oszillator (Bewegung parallel zur Defektachse) /100/.

Das Energieniveauschema einer solchen zweidimensionalen Bewegung des O-Atoms kann den Charakter einer Rotation, Libration oder Schwingung haben. Da jedoch die Energien der einzelnen Absorptionslinien grob betrachtet äquidistante Energieniveaus fordern, scheidet eine Rotationsbewegung aus. Ähnliches gilt auch für eine Librationsbewegung, d.h. eine Torsionsschwingung der Si-O-Si-Ebene, die im Si-Gitter sechs gleichberechtigte Gleichgewichtslagen hat, um eben diese. Hier sollte das Energieniveauschema infolge der Tunnelmöglichkeit zwischen den verschiedenen Lagen eine Feinstrukturaufspaltung zeigen /102/. Da eine solche nicht beobachtet wurde, schlossen Bosomworth et al. daraus, daß die Bewegung des O-Atoms im wesentlichen vibronischer Art ist und daß eine Wechselwirkung des O-Atoms mit den zweit- und drittnächsten Si-Nachbaratomen vernachlässigt werden kann. Es ist also möglich, eine axiale Symmetrie um die [111]-Achse für die Bewegung des O-Atoms anzunehmen. Die beobachteten Absorptionslinien können sehr gut mit dem Potential eines gestörten zweidimensionalen harmonischen Oszillators erklärt werden:

$$V = \frac{1}{2} M \omega_o^2 r^2 + A \exp(-q \gamma r^2) \quad (39)$$

Dabei ist M die Masse des O-Atoms, r sein radialer Abstand von der Si-Si-Verbindungsachse und $\gamma = M \omega_o / \hbar$. A , ω_o und q sind phänomenologische Konstanten, die dem Experiment zu entnehmen sind. Der Potentialberg bei $r = 0$ wurde empirisch eingeführt, um die außeraxiale Einbaulage des O-Atoms zu berücksichtigen. Die beobachteten Absorptionslinien in Abb. 39 b ergeben sich für $q = 1$, $A = 80 \text{ cm}^{-1}$ und $\omega_o = 49 \text{ cm}^{-1}$ (Einsetzen der

Isotopenmasse von ^{18}O liefert auch die in dem mit ^{18}O dotierten Kristall beobachtete zusätzliche Linie in Abb. 39 c). Das Potential und die daraus berechneten Energieniveaus zeigt Abb. 40.

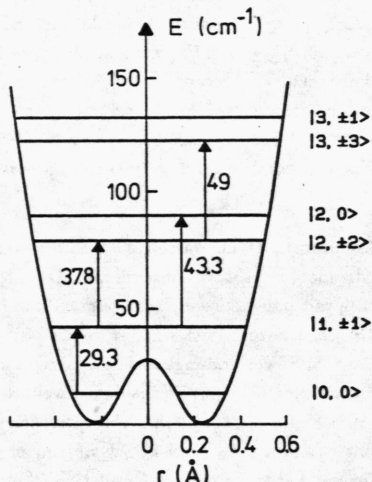


Abb. 40 Potential und daraus berechnete Energieniveaus für die Bewegung des Sauerstoffatoms senkrecht zur Si-Si-Verbindungsachse nach /100,103/.

Eine Kennzeichnung der Niveaus ist möglich mit den Quantenzahlen $|v, l\rangle$ des ungestörten zweidimensionalen harmonischen Oszillators, wobei v die Schwingungsquantenzahl angibt und l den Winkelanteil der Wellenfunktion beschreibt. Infolge der Störung des harmonischen Potentials durch den Potentialwall bei $r = 0$ erfolgt eine Aufspaltung der $(v + 1)$ -fach entarteten Energieniveaus des zweidimensionalen harmonischen Oszillators nach $|l|$: aus dem dreifach entarteten $v = 2$ Niveau entsteht ein Singulett und ein Dublett, aus dem vierfach entarteten $v = 3$ Niveau 2 Dubletts. (Zusätzlich würde die Gruppentheorie eine Aufspaltung des Niveaus $|3, \pm 3\rangle$ infolge der D_{3d} -Symmetrie des oben angegebenen Potentials im Si-Gitter erfordern. Daß eine solche in den Spektren nicht beobachtet wird, bestätigt die Annahme, daß keine oder nur eine sehr geringe Wechselwirkung des O-Atoms mit den zweit- und drittnächsten Si-Nachbaratomen stattfindet /100/.)

Mit den Auswahlregeln $\Delta v = 1$, $\Delta l = \pm 1$ für die Anregung einer σ -Oszillation erhält man die beobachteten Absorptionslinien aus Abb. 39. Im Bild der Eigenschwingungen eines freien Si_2O -Moleküls würde die eben diskutierte Bewegung des Sauerstoffatoms der

ν_2 -Biegeschwingung entsprechen. Damit besteht wieder eine qualitative Übereinstimmung mit gasförmigem Disiloxan, wo ebenfalls solche niederenergetischen Anregungen (in der Größenordnung von 60 cm^{-1}) beobachtet wurden /104/. Da alle Übergänge vom ersten angeregten Zustand $|1, \pm 1\rangle$ in weiter darüberliegende Zustände höhere Photonenenergien erfordern als der Übergang $|0, 0\rangle \rightarrow |1, \pm 1\rangle$ (vgl. Abb. 39 b), kann das Niveauschema für die Sauerstoffschwingung bezüglich der Absorption von FIR-Strahlung der Frequenz 29.3 cm^{-1} als ein Zweiniveausystem betrachtet werden. Es sei noch bemerkt, daß eigene Absorptionsmessungen mit Hilfe des CH_3F -Raman-Lasers eine unwesentlich geringere Absorptionsfrequenz von $(29.2 \pm 0.05) \text{ cm}^{-1}$ ergaben. Dieser Wert soll im weiteren Verlauf dieses Kapitels verwendet werden.

Die Ergebnisse von Bosomworth et al. lassen auch sofort eine Interpretation der $8 \mu\text{m}$ - und $9 \mu\text{m}$ -Absorptionsbande zu: Absorptionsexperimente unter Druckausübung in $[111]$ -Richtung bei Bestrahlung mit polarisiertem Licht zeigten, daß beide Absorptionsbande der Anregung von π -Oszillationen zuzuordnen sind. Eine π -Oszillation, d.h. eine Bewegung des O-Atoms parallel zur Si-Si-Verbindungsachse, entspricht der zu Beginn diskutierten antisymmetrischen Streckschwingung ν_3 . Wie aus dem erweiterten Niveauschema in Abb. 41 ersichtlich, resultieren die $8 \mu\text{m}$ - bzw. $9 \mu\text{m}$ -Absorptionsbande aus der Anregung von Kombinationsschwingungen aus den ν_2 - und ν_3 -Schwingungen unter Berücksichtigung der Auswahlregel $\Delta l = 0$ für die Anregung einer π -Oszillation. So lassen sich beispielsweise die Absorptionslinien bei 1136 cm^{-1} bzw. 1203 cm^{-1} als folgende Übergänge interpretieren:

$$1136 \text{ cm}^{-1} : |\nu_3, v, l\rangle = |0, 0, 0\rangle \rightarrow |1, 0, 0\rangle$$

$$1203 \text{ cm}^{-1} : |\nu_3, v, l\rangle = |0, 0, 0\rangle \rightarrow |1, 2, 0\rangle$$

Der ebenfalls in Abb. 41 eingezeichnete Übergang bei $19 \mu\text{m}$ (517 cm^{-1}) wird von Stavola /105/ anhand von Messungen des druckinduzierten Dichroismus in Si:O mit der Anregung der symmetrischen Streckschwingung ν_1 (bei Disiloxan: 606 cm^{-1}) verknüpft. Da ihre Frequenz, wie bereits erwähnt, unterhalb der Ramanfrequenz des ungestörten Siliziumgitters liegt, darf sie jedoch nicht auf der Basis eines freien Si_2O -Moleküls betrachtet werden, sondern muß vielmehr im Sinne einer resonanten Gittermode interpretiert werden /96,100/ (siehe folgender Abschnitt).

Als Abschluß sei noch bemerkt, daß der Absorptionskoeffizient α im Zentrum der $9 \mu\text{m}$ -Bande (gemessen bei $T = 300 \text{ K}$) üblicherweise als Maß für die Konzentration des interstitiellen Sauerstoffs $[\text{O}_I]$ in Silizium verwendet wird. Die aktuellst ermittelte Umrechnung ist:

$$[\text{O}_I] = 3.03 \cdot 10^{17} \cdot \alpha \quad [\text{cm}^{-3}] \quad /106/ \quad (40)$$

Hierbei ist α in cm^{-1} einzusetzen.

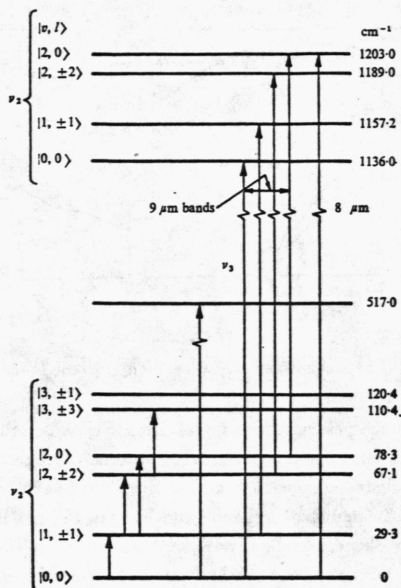


Abb. 41 Niveauschema für Sauerstoffschwingungen in Silizium mit den erlaubten Übergängen im mittleren und fernen Infrarot /100/.

1.4 Resonante Streuung akustischer Phononen an interstitiellen Sauerstoffatomen in Silizium

In der bisherigen Diskussion der Eigenschaften des Systems Si:O wurde nur eine lokale Oszillation des Sauerstoffatoms in einem als starr angenommenen Siliziumgitter betrachtet. Es stellt sich die Frage, was nun letztendlich die endliche Breite der Ferninfrarotabsorptionslinie verursacht. Da die Grundswingungsfrequenz von interstitiellem Sauerstoff $\nu_o = 29.2 \text{ cm}^{-1}$ ($= 876 \text{ GHz}$) im Frequenzbereich der akustischen Phononen in Silizium liegt (siehe die Phonondispersionskurven von Silizium in Abb. 42), ist die Vorstellung der isolierten Bewegung des Sauerstoffatoms im Siliziumgitter jedoch nicht haltbar.

Das dynamische Verhalten des Systems Si:O bei ν_o ist vielmehr das einer resonanten Bandmode. Dieser Begriff sei kurz qualitativ erläutert: Der Einbau eines Punktdefekts in einen ansonsten idealen Kristall führt zum Verlust der Translationssymmetrie und damit zu einer

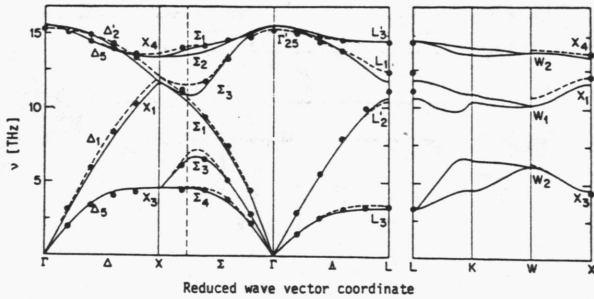


Abb. 42 Phononendispersionskurven von Silizium /107/.

Veränderung des Schwingungsspektrums des ungestörten Kristalls. Punktdefekte können dabei substitutionelle Störstellen (Isotope bzw. Fremdatome an regulären Gitterplätzen), Fehlstellen, interstitielle Störstellen oder auch molekulare Defekte sein, die zusätzliche Freiheitsgrade durch innere Schwingungen besitzen. Macht man für die Bewegung des κ -ten Atoms ($\kappa = 1 \dots r$) der l -ten Elementarzelle ($l = 1 \dots N$) in Richtung α ($\alpha = 1 \dots 3$) den Ansatz $u_\alpha(\kappa l; t) = u_\alpha(\kappa l)e^{i\omega t}$, so wird das Normalschwingungsspektrum eines gestörten Kristalls in der harmonischen Näherung beschrieben durch die Eigenwertgleichung

$$(\mathbf{D} - \omega^2 \mathbf{1} + \mathbf{\Gamma}) \mathbf{u} = \mathbf{0} \quad (41)$$

Dabei ist \mathbf{D} die $3rN$ -dimensionale dynamische Matrix des ungestörten Kristalls, ausgedrückt durch die atomaren Massen und Kraftkonstanten $D_{\alpha\beta}(l\kappa; l'\kappa') = \frac{1}{\sqrt{M_\kappa M_{\kappa'}}} \phi_{\alpha\beta}(l\kappa; l'\kappa')$. $\mathbf{\Gamma}$ ist die Störmatrix, die die Änderung der atomaren Massen und Kraftkonstanten bei Einfügen des Defekts beschreibt, d.h. $\Gamma_{\alpha\beta}(l\kappa; l'\kappa') = \frac{1}{\sqrt{M_\kappa M_{\kappa'}}} \Delta \phi_{\alpha\beta}(l\kappa; l'\kappa') - \omega^2 \frac{\Delta M_\kappa}{M_\kappa} \delta_{l\kappa; l'\kappa'}$, und die i.a. nur für Orte ($l\kappa$) in unmittelbarer Nähe des Störatoms von Null verschieden ist.

Im Gegensatz zum ungestörten Kristall (gegeben durch $\mathbf{\Gamma} = \mathbf{0}$), wo die Eigenmoden bekanntlich ebene Wellen sind, charakterisiert durch Wellenvektor \vec{k} und Frequenz $\omega_p(\vec{k})$ (p bezeichne die verschiedenen Zweige der Phononendispersionskurven), sind die Eigenmoden des gestörten Kristalls keine ebenen Wellen mehr. Wenn auch das resultierende Frequenzspektrum nahezu den gleichen Frequenzbereich und die gleiche Zustandsdichte besitzt wie das des ungestörten Kristalls, können doch (abhängig von der Störmatrix $\mathbf{\Gamma}$) einige zusätzliche Phänomene auftreten: so sind Eigenmoden möglich, deren Frequenz oberhalb der erlaubten Frequenzen im ungestörten Kristall liegt. Sie werden als lokalisierte Moden bezeichnet

und sind dadurch charakterisiert, daß das Defektatom mit großer Amplitude schwingt, die Amplitude der benachbarten Atome jedoch annähernd exponentiell mit dem Abstand vom Störatom abnimmt; eine Wellenausbreitung im Gitter ist bei dieser Frequenz nicht möglich. Ein Beispiel dafür ist die antisymmetrische Streckschwingung ν_3 des interstitiellen Sauerstoffs in Silizium, die ja mit ihrer Frequenz $\nu_3 = 1136 \text{ cm}^{-1}$ weit oberhalb der Ramanfrequenz ($519 \text{ cm}^{-1} / 99/$) des ungestörten Si-Gitters liegt. Ähnliches gilt für Eigenmoden, deren Frequenz in eine Lücke (falls vorhanden!) im Frequenzspektrum des ungestörten Kristalls fällt (Gapmoden). Sogenannte resonante Bandmoden treten auf, wenn die Schwingungsfrequenz ν_o des Defektes in einem als starr betrachteten Gitter (auch Einstein-Oszillationsfrequenz genannt) innerhalb des erlaubten Frequenzbereichs des ungestörten Gitters liegt. Sie bezeichnen jedoch keine speziellen Eigenmoden des gestörten Kristalls, sondern vielmehr eine gemeinsame Eigenschaft vieler Eigenmoden: für alle Eigenmoden innerhalb einer Breite Γ um die Frequenz ν_o (die durch die dynamische Antwort des Gitters auf die Defektschwingung etwas verschoben werden kann) gilt, daß die Amplitude des Defektatoms bzw. seiner unmittelbaren Nachbaratome, mit denen es wechselwirkt, sehr viel größer ist als die der restlichen Gitteratome, die jedoch (im Gegensatz zu den lokalisierten Moden) im gesamten Kristall von Null verschieden ist. Daher spricht man auch von einer pseudo-lokalisierten Mode. Eine resonante Bandmode wird also charakterisiert durch eine Resonanz für die Amplitude des Störatoms bzw. seiner unmittelbaren Nachbaratome für Schwingungsfrequenzen um ν_o .

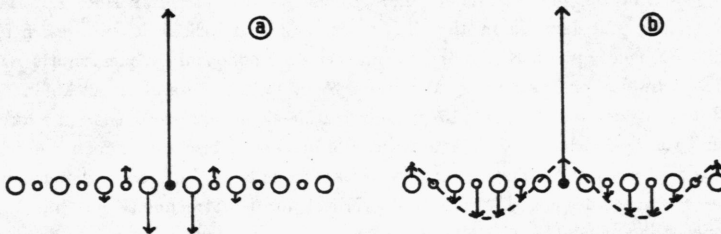


Abb. 43 Schematische Darstellung der Eigenmoden einer zweiatomigen linearen Kette in der Umgebung der Störstelle nach /108/: (a) lokalisierte Mode, (b) resonante Bandmode.

Der Einbau von Fremdatomen, auch wenn sie im freien Zustand elektrisch neutral sind, kann zu einer Veränderung der Dichteverteilung für die Valenzelektronen gegenüber der des ungestörten Kristalls führen; ist beispielsweise die Ladungsverteilung in der Umgebung des Störatoms nach dessen Einbau ausgedehnter als vorher, läßt sich das bildlich durch die

Einführung einer effektiven positiven Ladung des Störatoms und kompensierenden negativen Ladungen der umgebenden Atome ausdrücken /109/. Die zur resonanten Bandmode beitragenden Eigenmoden können somit je nach ihrer Symmetrie ein nichtverschwindendes Dipolmoment besitzen und damit infrarotaktiv sein. In diesem Fall kann die resonante Bandmode durch Absorption von FIR-Strahlung angeregt und nachgewiesen werden.

Eine ausführliche Übersicht über experimentelle Ergebnisse und Untersuchungsmethoden von lokalisierten und resonanten Moden in gestörten Kristallen, insbesondere in dotierten Alkalihalogenidkristallen, findet man in den Übersichtsartikeln von Barker et al. /110/ und Bilz et al. /111/, detaillierte theoretische Abhandlungen in den Artikeln von Maradudin /109/, Klein /112/ und Takeno /113/. Aus den genannten Arbeiten seien kurz einige wichtige Resultate, die für die Diskussion der experimentellen Ergebnisse dieser Arbeit nötig sind, diskutiert.

So erhält man eine spezielle Beschreibungsmöglichkeit für die Lösungen der Eigenwertgleichung (41) mit Hilfe der Streutheorie, indem man die Eigenmoden des gestörten Kristalls als Summe einer einlaufenden Welle, die eine Eigenmode des ungestörten Kristalls ist, und einer am Störatom (elastisch) gestreuten Welle ansetzt /109,114/. Eine resonante Bandmode entspricht in dieser Beschreibung einer Resonanz der Amplitude der gestreuten Welle. Somit ist der Streuquerschnitt $\sigma = \sigma(\nu)$ bzw. die Streurrate $\tau_{res}^{-1}(\nu) = N \bar{v} \sigma(\nu)$ (N = Konzentration der Störstellen, \bar{v} = mittlere Schallgeschwindigkeit) für die Phononen des ungestörten Kristalls eine geeignete Größe für die Beschreibung der dynamischen Eigenschaften eines gestörten Kristalls. ($\tau_{res}(\nu)$ gibt dann die mittlere Zeit zwischen 2 (elastischen) Streuprozessen für Phononen der Frequenz ν an). Konkrete Berechnungen des Phononenstreuquerschnitts $\sigma(\nu)$ wurden beispielsweise von Takeno bzw. Maradudin /109,113/ für isotopische Defekte in kubischen Bravaisgittern, von Wagner /115/ für substitutionelle molekulare Defekte in solchen Gittern und von Dederichs /116/ für interstitielle Störstellen kubischer Symmetrie durchgeführt. Sie alle zeigen, daß, wenn eine resonante Bandmode bei ν_0 im Frequenzbereich der akustischen Phononen des ungestörten Kristalls vorliegt, für den Streuquerschnitt (im Falle der isotropen Näherung $\omega_p(\vec{k}) = \bar{v} |\vec{k}|$) gilt:

$$\tau_{res}^{-1}(\nu) \sim \frac{\Gamma^2 \nu^2}{(\nu^2 - \nu_0^2)^2 + \Gamma^2 \nu^2} \simeq \frac{\Gamma^2/4}{(\nu - \nu_0)^2 + \Gamma^2/4} \quad (42)$$

Die letzte Näherung, gültig für den Fall $\Gamma \ll \nu_0$, ergibt eine Lorentzsche Linienform mit der Halbwertsbreite Γ für die Streurrate $\tau_{res}^{-1}(\nu)$. Daß der Absorptionskoeffizient für FIR-Strahlung (bei einer IR-aktiven resonanten Bandmode) die gleiche Frequenzabhängigkeit besitzt, zeigen Klein bzw. Maradudin in ihren Übersichtsartikeln /109,112/.

Dederichs und Zeller /116/ geben im Falle einer interstitiellen Störstelle kubischer Symmetrie

in einem einatomigen Gitter folgenden Wert für die Halbwertsbreite Γ an:

$$\frac{\Gamma}{\nu_o} = \frac{3}{16\pi^2} \nu_D^3 \frac{f_{eff}^{\frac{3}{2}}}{M \cdot M_{eff}^{\frac{1}{2}}} \quad \text{bzw.} \quad \Gamma = \frac{3}{2} \pi \frac{M_{eff}}{M} \frac{\nu_o^4}{\nu_D^3} \quad (43)$$

mit $\nu_o = \frac{1}{2\pi} \sqrt{\frac{f_{eff}}{M_{eff}}}$

Dabei ist M die Masse eines Gitteratoms und ν_D die Debyefrequenz des ungestörten Gitters. f_{eff} ist eine effektive Federkonstante, die die Bindung des interstitiellen Defekts an die umgebenden Gitteratome beschreibt, M_{eff} eine effektive Masse, die berücksichtigt, daß mit der Defektschwingung auch ein Mitschwingen der Umgebung verbunden ist (sie kann größer oder kleiner als die reale Defektmasse sein). Berechnungen für die effektiven Größen finden sich in /116/. Diese Ergebnisse zeigen, daß die Resonanz um so schärfer ist, je kleiner die effektive Bindung des Defekts ans Gitter ist bzw. je schwerer die effektive Masse ist, und damit je geringer die Resonanzfrequenz liegt.

Setzt man in Formel (43) die Werte für den interstitiellen Sauerstoff in Silizium ein, auch wenn das Sauerstoffatom an keinem Gitterplatz mit kubischer Symmetrie sitzt, und nimmt als effektive Masse die reale Masse des Sauerstoffatoms, erhält man $\Gamma \approx 0.7$ GHz. (Silizium hat die Debyefrequenz $\nu_D = 13.54$ THz /99/.) Ähnliches erhielt Brice /114/ bei numerischen Berechnungen von Γ für interstitielle Störstellen kubischer Symmetrie in Silizium. Direkte Rechnungen für den interstitiellen Sauerstoff in Silizium wurden bisher nicht durchgeführt.

Für den Wert der Streurate im Linienzentrum gilt nach /109/

$$\tau_{res}^{-1}(\nu_o) = \frac{4N}{D(\nu_o)} \quad (44)$$

N ist die Defektkonzentration (sie sei klein genug, um eine Wechselwirkung der Defekte untereinander ausschließen zu können) und $D(\nu_o)$ die Zustandsdichte des ungestörten Kristalls bei der Resonanzfrequenz. Das gleiche Ergebnis erhält man auch aus einer Arbeit von Levinson /117/, der die Resonanzstreuung akustischer Phononen an einem elektronischen Zweiniveausystem (z.B. an dem $(\bar{E}, 2\bar{A})$ -System angeregter Cr^{3+} -Ionen in Rubin /118/) störungstheoretisch behandelt.

Zusammenfassend gelangt man zu folgender Interpretation der resonanten Bandmode bei 29.2 cm^{-1} im System Si:O: Sie läßt sich als eine (durch Ferninfrarotstrahlung anregbare) Einsteinoszillation des Sauerstoffatoms betrachten, die gekoppelt ist (und im Gleichgewicht steht) mit dem Schwingungssystem des ungestörten Siliziumkristalls infolge Emission und Reabsorption resonanter Phononen, d.h. von Phononen mit der gleichen Energie wie die

Sauerstoffschwingung. Dies entspricht der oben diskutierten Resonanzstreuung. (Unter dem Begriff Gleichgewicht mit dem Schwingungssystem des ungestörten Kristalls wird hier nicht das thermische Gleichgewicht verstanden, sondern nur das partielle Gleichgewicht mit den Phononen, die Frequenzen innerhalb der erwähnten Resonanzlinie besitzen. Eine Anregung der resonanten Bandmode bedeutet dann durchaus einen Nichtgleichgewichtszustand der resonanten Phononen bezüglich des thermischen Gleichgewichts!) Infolge der Lorentzischen Linienform für den Phononenstreuquerschnitt bzw. die Ferninfrarotabsorption kann man charakteristische Zeiten T_1 (Lebensdauer der Sauerstoffschwingung gegenüber der Emission eines resonanten Phonons) bzw. T_2 (Phasenrelaxationszeit für die Sauerstoffschwingung infolge der Streuung resonanter Phononen) definieren über die üblichen Beziehungen /119/

$$T_1 = \frac{1}{2\pi\Gamma} \quad \text{und} \quad T_2 = 2T_1, \quad (45)$$

die den Energieaustausch zwischen den Sauerstoff-Oszillationen und dem Gitterschwingungssystem des Siliziumkristalls beschreiben. Dabei ist vorausgesetzt, daß keine Energie aus dem gekoppelten System abfließt. Die Bestimmung obiger Werte aus den Ferninfrarotabsorptionsmessungen ist Gegenstand des folgenden Abschnitts.

II.2. Transmissionsspektren von Sauerstoff-dotiertem Silizium im fernen Infrarot

Die in diesem Abschnitt vorgestellten Messungen der Transmission von FIR-Strahlung an Sauerstoff-dotierten Siliziumproben sollen zum einen die Möglichkeit demonstrieren, das entwickelte abstimmbare FIR-Lasersystem für die FIR-Spektroskopie zu nutzen; zum anderen sollen aus ihnen Informationen über Form und Breite der FIR-Absorptionslinie der resonanten Bandmode gewonnen werden.

Meßmethode für die folgenden Transmissionsspektren war die in Abschnitt I.3.2 beschriebene Methode der Energietransmissionsmessung. Es wurde dabei darauf geachtet, daß die auf den Kristall auftreffenden Pulsenergien des FIR-Lasers stets unter $25 \mu\text{J}$ blieben; sie waren damit - wie im folgenden Abschnitt gezeigt - genügend gering, um Nichtlinearitäten des Absorptionskoeffizienten vernachlässigen zu können. Bei den Messungen befand sich die Probe in suprafluidem Helium bei einer Temperatur $T = 1.8 \text{ K}$.

Die Punkte in Abb. 44 zeigen die gemessenen Transmissionswerte einer Siliziumprobe mit einer Sauerstoffkonzentration $[\text{O}_i] = 18 \cdot 10^{17} \text{ cm}^{-3}$ bei einer Frequenzvariation der Strahlung des CH_3F -Raman-Lasers im Bereich von 25 cm^{-1} bis 34 cm^{-1} . Die senkrechten Striche geben die mittlere quadratische Abweichung bei einer Mittelung über 5 Laserpulse an. Die Fehler in den Frequenzwerten betragen etwa $\pm 0.05 \text{ cm}^{-1}$ (vgl. Abschnitt I.2.3.1). Das Spektrum, das

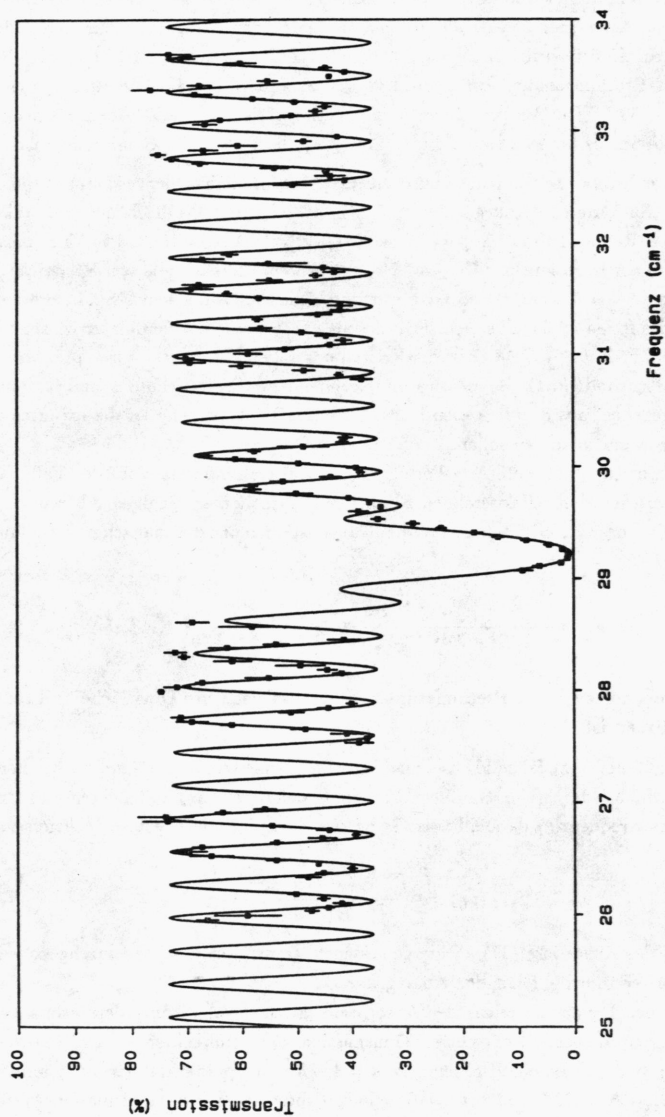


Abb.44 Transmissionsspektrum einer 5 mm dicken Siliziumprobe mit einer Sauerstoffkonzentration $[O] = 18 \cdot 10^{17} \text{ cm}^{-3}$. Die durchgezogene Linie ist das Ergebnis einer Datenanpassung an Gl.(47).

die einzelnen Abstimmintervalle des CH_3F -Raman-Lasers widerspiegelt, ist geprägt durch die Vielstrahlinterferenz der Laserstrahlung, die an den beiden parallelen Oberflächen der Siliziumprobe reflektiert wird. Denn da Silizium bei $T = 1.5 \text{ K}$ für FIR-Strahlung bei 30 cm^{-1} einen Brechungsindex von $n = 3.38$ besitzt [120], beträgt die Reflexion pro Oberfläche $R = 30\%$. Der Vielstrahlinterferenz überlagert ist die starke Absorptionslinie mit der Zentralfrequenz bei 29.2 cm^{-1} , die die Anregung der resonanten Bandmode anzeigt.

Daß eine einfache Entnahme von Form und Breite der Absorptionslinie der resonanten Bandmode aus dem Spektrum in Abb.44 nicht möglich ist, zeigt ein Vergleich der spektralen Breite des CH_3F -Raman-Lasers ($\Delta\nu_L = (4 \pm 0.5) \text{ GHz}$; vgl. Abschnitt I.2.3.1), der erwarteten Breite Γ der Absorptionslinie ($\Gamma \leq 6 \text{ GHz}$; vgl. Abschnitt II.1.3: die von Bosomworth angegebene Breite von $0.2 \text{ cm}^{-1} = 6 \text{ GHz}$ war durch die Auflösung seines Spektrometers gegeben) und des freien Spektralbereichs FSR des als Fabry-Perot-Interferometer wirkenden Siliziumkristalls ($\text{FSR} = \frac{c}{2nd} = 8.8 \text{ GHz}$; der Kristall war auf eine Dicke $d = 5 \text{ mm}$ präpariert; c ist die Lichtgeschwindigkeit). Daher wird im folgenden eine theoretische Beschreibung der Transmission gegeben, aus der die wesentlichen Daten durch Anpassung an die experimentellen Punkte gewonnen werden sollen:

Modifiziert man die wohlbekannte Airy-Formel für die Vielstrahlinterferenz für den Fall, daß das Medium zwischen den reflektierenden Flächen einen frequenzabhängigen Absorptionskoeffizienten $\alpha(\nu)$ besitzt, erhält man für die Transmission monochromatischer Strahlung der Frequenz ν

$$T(\nu) = \frac{e^{-\alpha(\nu)d} (1 - R)^2}{1 + R^2 e^{-2\alpha(\nu)d} - 2R e^{-\alpha(\nu)d} \cos\left(\frac{4\pi n d \nu}{c}\right)}, \quad (46)$$

wobei d die Probendicke, n der Brechungsindex, R die Reflexion pro Oberfläche und c die Lichtgeschwindigkeit ist.

Ist $E(\nu) d\nu$ die auf die Probe fallende Laserenergie im Frequenzintervall $[\nu, \nu + d\nu]$, wobei $E(\nu)$ nach Abschnitt I.2.3.1 durch eine Gaußsche Verteilung der Breite $\Delta\nu_L$ mit Zentrum an der eingestellten Laserfrequenz ν_L darstellbar ist, so beträgt die gemessene effektive Transmission $T_{eff}(\nu_L)$:

$$T_{eff}(\nu_L) = \frac{\int_0^\infty E(\nu) T(\nu) d\nu}{\int_0^\infty E(\nu) d\nu} \quad (47)$$

Die Anpassung des Ausdrucks (47) an die gemessenen Transmissionswerte (durchgezogene Linie in Abb.44) erlaubt die folgenden Aussagen:

- a) Außerhalb des Frequenzbereichs der Absorption durch die Anregung der resonanten Bandmode reproduziert die berechnete Transmissionskurve (für eine spektrale Breite des Lasers $\Delta\nu_L = 3.6 \text{ GHz}$, eine Probendicke $d = 4.983 \text{ mm}$ - präpariert war auf 5 mm ! -, eine Reflexion $R = 30\%$ und $\alpha(\nu) = 0$) sehr gut die gemessenen Transmissionswerte.

Damit wurde bereits auf eine sehr einfach nachprüfbare Weise die Verwendbarkeit des CH_3F -Raman-Lasers (trotz seiner strukturierten Emissionscharakteristik) für absorptionspektroskopische Untersuchungen in seinem Emissionsbereich gezeigt.

- b) Für die Beschreibung der FIR-Absorption durch die Anregung der resonanten Bandmode wurde gemäß den Betrachtungen des vorhergehenden Abschnitts eine Lorentzsche Linienform für den Absorptionskoeffizienten in Gl.(46) angenommen:

$$\alpha(\nu) = \alpha_0 \frac{\Gamma^2/4}{(\nu - \nu_0)^2 + \Gamma^2/4} \quad , \quad (48)$$

wobei $\nu_0 = 29.2 \text{ cm}^{-1}$ die beobachtete Zentralfrequenz der Absorptionslinie war. Die Anpassung lieferte für den Absorptionskoeffizienten im Linienzentrum $\alpha_0 = 18.5 \text{ cm}^{-1}$ und für die Halbwertsbreite $\Gamma = 0.13 \text{ cm}^{-1}$ bzw. 4 GHz. Eine Anpassung mittels einer Gaußschen Linienform für $\alpha(\nu)$ ergab ein deutlich schlechteres Ergebnis als die in Abb. 44 gezeigte Anpassung. Analoge Messungen wurden auch an Siliziumproben mit anderen Sauerstoffkonzentrationen und Probendicken durchgeführt. (Ein Beispiel für einen 1.3 mm dicken Siliziumkristall mit einer Sauerstoffkonzentration $[\text{O}_I] = 13 \cdot 10^{17} \text{ cm}^{-3}$ ist in Abb. 45 dargestellt. Dort beträgt $\alpha_0 = 12 \text{ cm}^{-1}$ und $\Gamma = 0.13 \text{ cm}^{-1}$). In allen Fällen erwies sich eine Anpassung mit Lorentzscher Linienform des Absorptionskoeffizienten als optimal, die Linienbreite Γ lag stets im Bereich $\Gamma = (0.13 \pm 0.01) \text{ cm}^{-1}$. Außerdem ergab sich eine lineare Abhängigkeit des Absorptionskoeffizienten α_0 von der Konzentration der Sauerstoffatome, was zeigt, daß letztere klein genug ist, um jedes Sauerstoffatom als isolierten Punktdefekt betrachten zu können. Aus der Halbwertsbreite Γ erhält man (vgl. Abschnitt II.1.4) für die Lebensdauer T_1 der Sauerstoffschwingungen gegenüber Emission resonanter Phononen $T_1 \approx 40 \text{ psec}$ bzw. für die Phasenrelaxationszeit $T_2 = 2T_1 = 80 \text{ psec}$. Diese Werte zeigen, daß während eines FIR-Laserpulses, dessen Halbwertsbreite typisch 70 nsec (vgl. Abschnitt I.2.3.2) beträgt, stets von einem (partiellen) Energiegleichgewicht zwischen den Sauerstoffschwingungen und dem Gitterschwingungssystem des Siliziumkristalls ausgegangen werden kann; denn auch wenn die FIR-Strahlung nur eine Anregung der Sauerstoffschwingungen bewirkt, erfolgt doch der Energieaustausch zwischen den beiden Systemen nach obigen Werten im Subnanosekundenbereich.

Es sei noch erwähnt, daß die gemessenen geringen Transmissionswerte im Absorptionszentrum ($T_{eff} < 1\%$) die in Abschnitt I.2.4 angesprochene Verwendbarkeit des CH_3F -Raman-Lasers zur Messung sehr kleiner Transmissionen auch an sehr kleinen Proben (der Durchmesser des Laserstrahls auf der Kristalloberfläche betrug weniger als 2 mm!) bestätigen.

Ein Demonstrationsbeispiel hochauflösender Spektroskopie mit Hilfe des CH_3F -Raman-Lasers ist in Abb. 45 zu finden. Wie bereits erwähnt, ist dort das Transmissionsspektrum

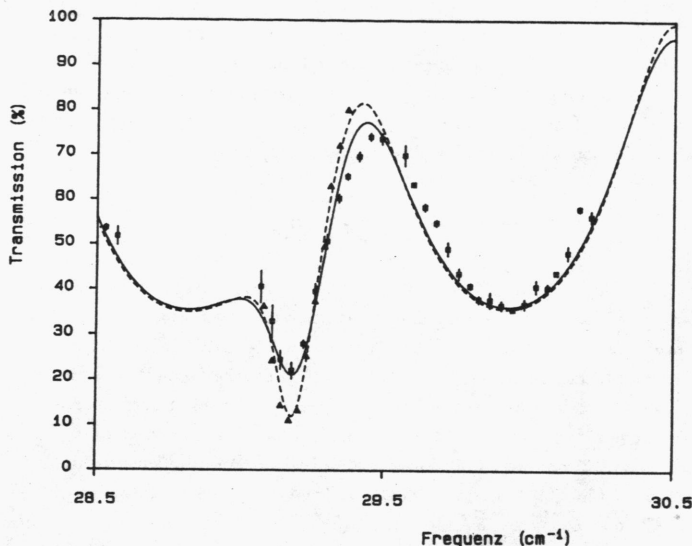


Abb. 45 Transmissionsspektrum einer Siliziumprobe mit einer Sauerstoffkonzentration $[O] = 13 \cdot 10^{17} \text{ cm}^{-3}$ für 2 verschiedene spektrale Breiten $\Delta\nu_L$ der Laserstrahlung: $\square \Delta\nu_L = 3.6 \text{ GHz}$; $\triangle \Delta\nu_L = 0.4 \text{ GHz}$. Die Kurven I, II sind die Ergebnisse der entsprechenden Datenanpassungen nach Gl.(47).

eines 1.3 mm dicken Siliziumkristalls mit einer Sauerstoffkonzentration $[O] = 13 \cdot 10^{17} \text{ cm}^{-3}$ auszugswise dargestellt. Die Quadrate sind die Transmissionswerte bei der gleichen spektralen Breite der FIR-Strahlung wie in Abb. 44, nämlich $\Delta\nu_L = 3.6 \text{ GHz}$; die durchgezogene Linie I zeigt die Anpassung nach Ausdruck (47) mit $\alpha_0 = 12 \text{ cm}^{-1}$ und $\Gamma = 0.13 \text{ cm}^{-1}$. Die Dreiecke stellen dagegen die Transmissionswerte dar, wenn die spektrale Breite der FIR-Strahlung mittels eines Fabry-Perot-Interferometers auf $\Delta\nu'_L \approx 0.4 \text{ GHz}$ reduziert wird (vgl. Abschnitt 1.2.3.1). Setzt man in Ausdruck (47) außer für $\Delta\nu_L$ die gleichen Parameter ein wie im Falle der durchgezogenen Kurve, erhält man die gestrichelte Kurve II. Die gute Übereinstimmung mit dem Experiment beweist die Richtigkeit der beschriebenen Auswertemethode auch im Falle der vollen spektralen Breite des CH_3F -Raman-Lasers. Im Gegensatz zu den Messungen von Bosomworth, der nur eine obere Grenze $\Gamma \leq 0.2 \text{ cm}^{-1}$ für die Halbwertsbreite der Absorptionslinie der resonanten Bandmode angeben konnte, folgt aus den

Ergebnissen dieser Arbeit: $\Gamma = (0.13 \pm 0.01) \text{ cm}^{-1}$ bzw. $(3.9 \pm 0.3) \text{ GHz}$. Dieser Wert liegt nahezu in der gleichen Größenordnung wie die Abschätzung von $\Gamma = 0.7 \text{ GHz}$ im Falle interstitieller Störstellen kubischer Symmetrie in Silizium im vorhergehenden Kapitel.

II.3. Nichtlineare Absorption von Ferninfrarotstrahlung in Sauerstoff-dotiertem Silizium

3.1 Experimentelle Ergebnisse

In diesem Abschnitt werden Messungen vorgestellt, die erstmals eine nichtlineare Absorption von FIR-Strahlung durch eine resonante Bandmode in gestörten Kristallen demonstrieren. Der Nachweis wurde erbracht durch Untersuchung der Transmission eines Sauerstoff-dotierten Siliziumkristalls für FIR-Strahlung hoher Pulsenergie mit Frequenzen in der unmittelbaren Umgebung der Resonanzfrequenz $\nu_0 = 29.2 \text{ cm}^{-1}$ der resonanten Bandmode. Der im Rahmen dieser Arbeit entwickelte CH_3F -Raman-Laser lieferte die dafür nötigen Voraussetzungen.

Bevor nun die Meßresultate gezeigt werden, seien zunächst kurz die für ihre Interpretation nötigen experimentellen Bedingungen angegeben:

- Als Probenkristall wurde der mit $18 \cdot 10^{17} \text{ cm}^{-3}$ Sauerstoffatomen dotierte Siliziumkristall verwendet, dessen Transmissionsspektrum anhand von Abb.44 diskutiert wurde.
- Der Kristall befand sich in suprafluidem Helium bei einer Temperatur $T = 1.8 \text{ K}$. Da dann $kT \ll h\nu_0$ (k = Boltzmannkonstante, h = Plancksches Wirkungsquantum) ist, kann jegliche thermische Anregung von Sauerstoffschwingungen vernachlässigt werden.
- Sämtliche Angaben der Pulsenergie der FIR-Strahlung kennzeichnen die auf die Probe treffende Energie und wurden nach Gl.(37) unter Verwendung des Korrekturfaktors $\gamma = 3.5$ bestimmt. (Eine Begründung für die Wahl dieses Korrekturfaktors wird im Laufe der folgenden Abschnitte gegeben.)
- Die Halbwertsbreite der durch ein Gaußsches Profil beschreibbaren Intensitätsverteilung der FIR-Strahlung (vgl. Abschnitt I.3.1) auf der Probenoberfläche betrug 1.7 mm . Eine Pulsenergie von $E = 100 \mu\text{J}$ ergibt somit im räumlichen Zentrum des Laserstrahls eine Energiedichte von $3 \frac{\text{mJ}}{\text{cm}^2}$ bzw. eine Spitzenintensität von $43 \frac{\text{kW}}{\text{cm}^2}$ bei einer Pulsdauer der FIR-Strahlung von 70 nsec .
- Die spektrale Breite der auf den Kristall treffenden Strahlung war die volle spektrale Breite des CH_3F -Raman-Lasers $\Delta\nu_L = 3.6 \text{ GHz}$.

In Abb.46 sind die Energietransmissionswerte für einzelne FIR-Pulse mit verschiedenen Pulsenergien bis zu $200 \mu\text{J}$ bei einer Laserfrequenz $\nu_L = 29.22 \text{ cm}^{-1}$ aufgetragen. Es zeigt

sich ein deutlicher Anstieg der Transmission T_{eff} mit der FIR-Pulsenergie: ausgehend von $T_{eff} = 0.2\%$ bei geringen Pulsenergien erreicht sie einen um den Faktor 10 höheren Wert von $T_{eff} = 2\%$ bei Pulsenergien von $200 \mu\text{J}$.

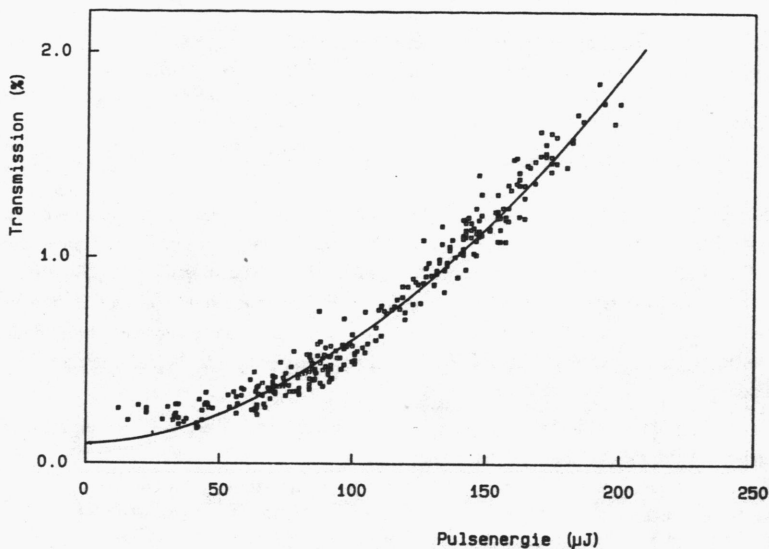


Abb. 46 Transmission einer 5 mm dicken Siliziumprobe mit einer Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ als Funktion der FIR-Pulsenergie für eine Laserfrequenz von 29.22 cm^{-1} ; die spektrale Breite des Detektors war auf 2.8 GHz eingeschränkt (siehe Text). Die durchgezogene Linie ist eine Modellrechnung gemäß Abschnitt II.3.2.4.

Die in Abb. 46 gezeigte Messung bedarf jedoch noch einer ergänzenden Bemerkung zu den experimentellen Bedingungen: Wie in Abschnitt I.2.3.1 erläutert, treten bei hohen Pulsenergien des CH_3F -Raman-Lasers im Emissionsspektrum des Lasers neben der (durch die Frequenz des Hochdruck- CO_2 -Lasers bestimmten) Frequenz der Ramanemission weitere Komponenten auf, die durch resonante, der Ramanemission nachfolgende Übergänge im Rotationsniveauschema eines CH_3F -Moleküls verursacht werden. Bei den in diesem Kapitel beschriebenen Experimenten wurde im Falle einer Laserpulsenergie größer als $250 \mu\text{J}$ - dies entspricht einer auf die Probe treffenden Energie von etwa $125 \mu\text{J}$ - neben der Frequenz der Ramanemission

auch Strahlung der Frequenz 28.5 cm^{-1} (Übergang $|1, J = 17\rangle \rightarrow |1, J = 16\rangle$ in CH_3F) beobachtet. Der Anteil dieser Strahlung in der gesamten Pulsenergie betrug zwar nur einige Prozent bei einer Laserpulsenergie von $400\text{ }\mu\text{J}$, aber da in den geschilderten Experimenten die Probe für Strahlung mit einer Frequenz in der Umgebung von ν_0 auch nur eine Transmission von einigen Prozent zeigte, konnte das Auftreten dieser zusätzlichen Strahlungskomponente, die ja im Siliziumkristall nicht absorbiert wurde, eine Nichtlinearität der Transmission vortäuschen. Aus diesem Grund wurde das in Abschnitt I.2.2 beschriebene Fabry-Perot-Interferometer (eingestellt auf die Frequenz der Ramanemission ν_L mit einer Bandbreite $\Delta\nu_{FP} = 2.8\text{ GHz}$) zwischen Probe und Transmissionsdetektor eingefügt, um die 28.5 cm^{-1} -Strahlung zu unterdrücken. (Ein Einfügen des Fabry-Perot-Interferometers zwischen Laser und Probe war in diesem Fall infolge der hohen Energieverluste der Laserstrahlung beim Durchgang durch das Interferometer [etwa 70% - vgl. Abschnitt I.2.2] nicht sinnvoll.) Die Strahlung wurde daher nur in einer spektralen Breite von 2.8 GHz detektiert, wenngleich auch ihre spektrale Breite beim Auftreffen auf den Kristall 3.6 GHz betrug.

Wie die Energietransmission als Funktion der auf die Probe fallenden FIR-Pulsenergie für den Fall aussieht, daß kein Fabry-Perot-Interferometer die 28.5 cm^{-1} -Strahlung unterdrückt, zeigt Abb.47 für drei verschiedene FIR-Frequenzen ($\nu_L = 29.22\text{ cm}^{-1}$, 29.25 cm^{-1} und 29.28 cm^{-1} ; alle Frequenzangaben erfolgten durch Vergleich der Transmission bei niedriger Pulsenergie mit dem Transmissionspektrum in Abb.44 und der daran angepaßten effektiven Transmission nach Gl.(47)). Auch hier sieht man eine deutliche Zunahme der Transmission mit der FIR-Pulsenergie. Die durchgezogenen Linien in Abb.46 und 47 stellen das (vorweggenommene) Ergebnis einer Modellrechnung dar, deren Inhalt Gegenstand der nächsten Kapitel ist. Während die gerechnete Transmissionskurve bei Abb.46 sehr gut mit den experimentellen Werten übereinstimmt, war bei den Experimenten ohne Interferometer in Abb.47 eine Beschreibung der Transmissionskurven nur für Pulsenergien unter $125\text{ }\mu\text{J}$ möglich; die Abweichung oberhalb dieser Energie ist eine Folge des Einsetzens der Emission der 28.5 cm^{-1} -Strahlung. Die Ursache der Streuung der Transmissionswerte in Abb.46 und Abb.47 liegt darin, daß auch bei gleichbleibender Laserpulsenergie von Puls zu Puls sich sowohl die spektrale Verteilung der Laserstrahlung ändert (nur über mehrere Pulse gemittelt ergibt sich eine annähernd Gaußsche Verteilung mit der Breite $\Delta\nu_L$ und dem Zentrum bei ν_L) als auch die zeitliche Struktur der Laserpulse, die, wie in den folgenden Abschnitten gezeigt wird, Einfluß auf die Energietransmission hat. Auch die geringen Schwankungen in der Ausbreitungsrichtung der Laserstrahlung durch die von Puls zu Puls verschiedene Modenausbreitung im Wellenleiterrohr des CH_3F -Raman-Lasers spielt nach Abschnitt I.3.1 eine Rolle.

Eine wichtige Ergänzung für die Analyse der nichtlinearen Energietransmission lieferten Messungen der zeitlichen Änderung der Transmission während eines Laserpulses hoher Energie. Bei diesen Messungen wurde die zeitliche Struktur des auf die Probe fallenden FIR-

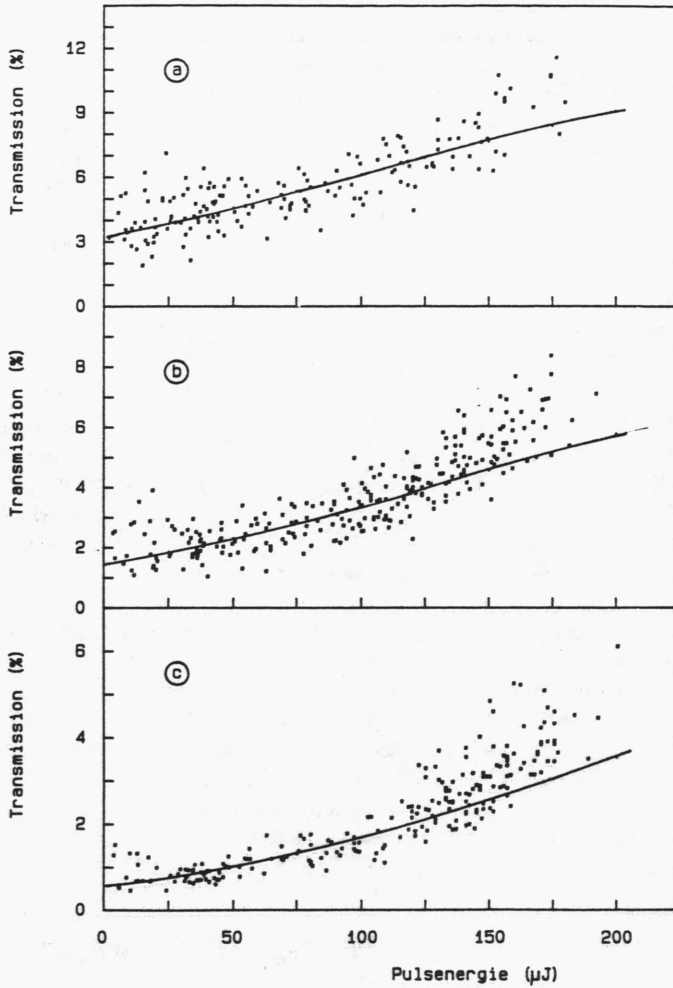


Abb. 47 Transmission einer 5 mm dicken Siliziumprobe mit einer Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ als Funktion der FIR-Pulsenergie für drei verschiedene Laserfrequenzen a) 29.28 cm^{-1} , b) 29.25 cm^{-1} , c) 29.22 cm^{-1} . Die spektrale Breite des Detektors war nicht begrenzt (siehe Text). Die durchgezogenen Linien sind das Ergebnis von Modellrechnungen gemäß Abschnitt II.3.2.4.

Laserpulses sowie des transmittierten Pulses simultan mittels 2 Schottkydioden gemäß der Methode in Abschnitt I.3.2 aufgezeichnet und daraus die Transmission bestimmt. Abb. 48 a) und b) zeigen als Beispiel die beiden Pulsstrukturen für einen FIR-Puls mit einer Einfallensenergie von $110 \mu\text{J}$ und einer Frequenz von 29.22 cm^{-1} . (Das Fabry-Perot-Interferometer wurde hier nicht verwendet.) Aufgetragen wurden in Abb. 48 die Intensitäten der beiden Pulse im räumlichen Zentrum der FIR-Strahlung; denn da der Antennenquerschnitt der Schottkydioden (hier näherungsweise als λ_L^2 angenommen, wobei $\lambda_L = 342 \mu\text{m}$ die Wellenlänge der FIR-Strahlung kennzeichnet) klein ist gegen die Halbwertsbreite von 1.7 mm der Querschnittsverteilung der Strahlung am Ort der Detektoren, ist die Intensität, nicht die Leistung die beobachtete Meßgröße; die Berechnung ihres Absolutwertes erfolgte aus der gemessenen Pulsenergie zusammen mit der gemessenen zeitlichen Pulsstruktur unter Berücksichtigung der Verteilung der Intensität der FIR-Strahlung innerhalb ihres Strahlquerschnitts (siehe Abschnitt I.3.1).

In Abb. 48 c) ist die resultierende Transmission dargestellt, die einen starken Anstieg in der ersten Hälfte des FIR-Pulses zeigt (durchgezogene Linie; die gestrichelte Linie ist das Ergebnis der erwähnten Modellrechnung). Das zum Intensitätsmaximum des einfallenden FIR-Pulses zeitlich versetzte Maximum der Transmission sowie ihr leichter Abfall in der zweiten Hälfte des Pulses deuten bereits an, daß eine hohe Anregung der resonanten Bandmode durch Ferninfrarotabsorption zwar für Zeiten länger als die Dauer des FIR-Pulses existiert, daß aber Dämpfungsmechanismen bereits während des Laserpulses wirksam sind.

3.2 Analyse der experimentellen Ergebnisse

Ziel dieses Kapitels ist, die (bereits angedeutete) Dämpfung der durch Absorption von FIR-Strahlung stark angeregten resonanten Bandmode genauer zu analysieren. So sollen anhand der beobachteten nichtlinearen Transmission der FIR-Strahlung (und einiger zusätzlicher Experimente, die noch geschildert werden) Rückschlüsse gezogen werden, wie und innerhalb welcher Zeit die durch die FIR-Absorption injizierte Energie aus dem (durch Emission und Reabsorption resonanter Phononen) gekoppelten System von Sauerstoffschwingungen und Gitterschwingungen des Siliziumkristalls abfließt. Zu diesem Zweck wird zunächst die Abhängigkeit der Transmission eines Sauerstoff-dotierten Siliziumkristalls von der FIR-Pulsenergie anhand eines einfachen Ratengleichungsmodells für den Fall berechnet, daß während des FIR-Pulses keine Energie aus dem gekoppelten System verlorengeht. Die Ergebnisse zeigen eine deutlich stärkere Zunahme der Transmission mit der FIR-Pulsenergie, als experimentell beobachtet wurde, was einen erneuten Beweis darstellt, daß eine Dämpfung der resonanten Bandmode bereits während des Laserpulses erfolgt. Mögliche Dämpfungsmechanismen werden diskutiert. Experimente zur Ausbreitung der resonanten Phononen, erzeugt durch Relaxation

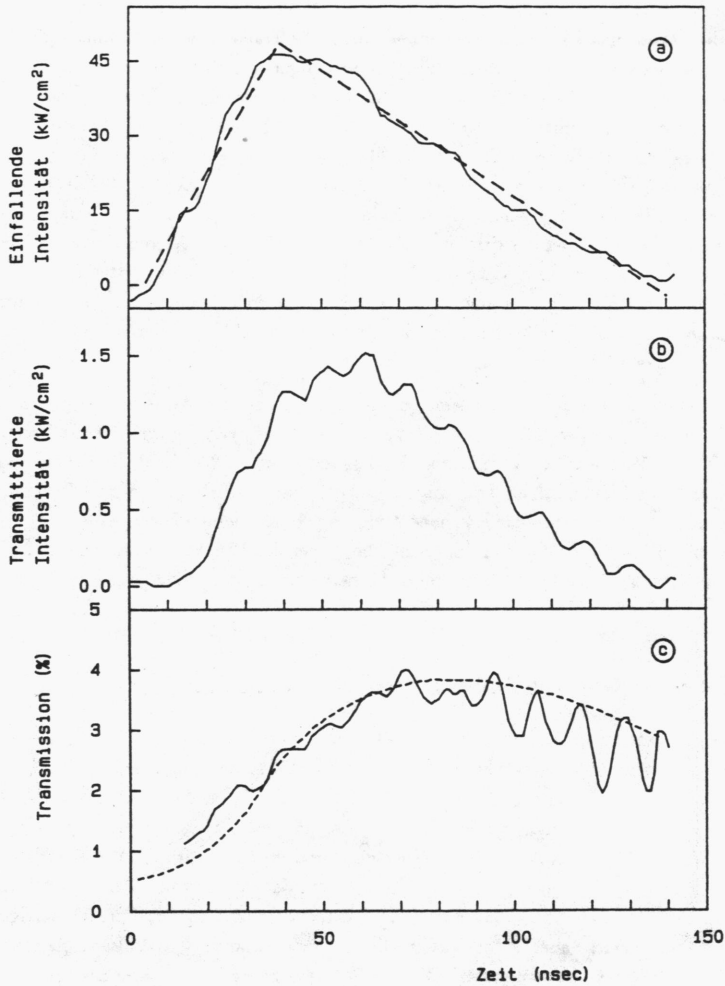


Abb. 48 Zeitliche Änderung der Transmission eines 5 mm dicken Siliziumkristalls mit einer Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ für einen FIR-Puls der Energie $110 \mu\text{J}$ (durchgezogene Kurve in c); die gestrichelte Kurve stellt das Ergebnis einer Modellrechnung nach Abschnitt II.3.2.4 dar). a) und b) zeigen die zeitliche Struktur des einfallenden und transmittierten FIR-Pulses. (Die gestrichelte Kurve in a) ist eine idealisierte zeitliche Struktur, die in der erwähnten Modellrechnung verwendet wird.)

der Sauerstoffschwingungen nach Anregung durch die FIR-Strahlung, werden beschrieben; sie ermöglichen den Schluß, daß eine sogenannte räumlich-spektrale Diffusion (siehe Abschnitt II.3.2.3) dieser Phononen den dominanten Dämpfungsmechanismus für die resonante Bandmode darstellt. Es wird gezeigt, daß eine Einbeziehung dieses Mechanismus in die Modellrechnung eine gute Übereinstimmung mit den experimentellen Resultaten liefert.

3.2.1 Modellrechnung für die Abhängigkeit der Transmission eines Sauerstoff-dotierten Siliziumkristalls von der FIR-Pulsenergie für den Fall vernachlässigbarer Dämpfung der resonanten Bandmode

Die grundlegenden Ideen zur Berechnung der Transmission eines Sauerstoff-dotierten Siliziumkristalls für gepulste FIR-Strahlung im Falle verschwindender Dämpfung der resonanten Bandmode sind im folgenden kurz skizziert:

- a) Gemäß den Betrachtungen in den Abschnitten II.1.3 und II.1.4 läßt sich das System der Sauerstoffschwingungen mit einem System von untereinander identischen Zweiniveausystemen identifizieren, deren Energieabstand durch die Resonanzfrequenz ν_0 der resonanten Bandmode gegeben ist und die homogen im Siliziumkristall verteilt sind. Ihre Konzentration ist gegeben durch die Konzentration N der Sauerstoffatome. Die Wechselwirkung dieser Zweiniveausysteme mit dem Schwingungssystem des Siliziumkristalls durch ständige Emission und Reabsorption resonanter Phononen, die nach II.1.4 die resonante Bandmode ausmacht, läßt sich durch folgende Rategleichung beschreiben (Eintreffen von FIR-Strahlung wird dabei noch nicht betrachtet):

$$\frac{dN_2(t)}{dt} = -\frac{N_2(t)}{T_1} + \int_0^{\infty} \frac{Z(\nu, t) d\nu}{\tau_{res}(\nu)} \quad (49)$$

Hierbei ist $N_2(t)$ die Konzentration angeregter Zweiniveausysteme und $Z(\nu, t) d\nu$ die Konzentration der Phononen mit Frequenzen im Intervall $[\nu, \nu + d\nu]$. T_1 gibt die Lebensdauer der angeregten Zweiniveausysteme gegenüber der spontanen Emission von Phononen an, $\tau_{res}(\nu)$ die mittlere Zeit zwischen 2 Reabsorptionsprozessen von Phononen der Frequenz ν . Der erste Term in Gl.(49) beschreibt somit die Erzeugung von Phononen durch spontane Relaxation von angeregten Zweiniveausystemen, der zweite Term die Reabsorption von Phononen durch im Grundzustand befindliche Zweiniveausysteme. Das Integral ist nötig, da die spontan emittierten Phononen eine Frequenzverteilung endlicher Breite besitzen (siehe Punkt 2).

Bei der Aufstellung von Gl.(49) ist angenommen, daß die Emission eines resonanten Phonons den einzigen Relaxationsprozeß für angeregte Zweiniveausysteme darstellt, und

daß sich das resonante Phonon unbeeinflußt ballistisch durch den Kristall bewegt, bis es nach der Zeit $\tau_{res}(\nu)$ stets wieder von einem Zweiniveausystem reabsorbiert wird. Es geht somit keine Energie aus dem gekoppelten System von Zweiniveausystemen und resonanten Phononen verloren: die resonante Bandmode ist ungedämpft. Damit kann man eine Gleichgewichtsbedingung angeben, die beschreibt, wie sich eine im gekoppelten System enthaltene Energie (die z.B. über FIR-Absorption durch die Zweiniveausysteme injiziert wurde) nach genügend langer Zeit auf die Sauerstoffschwingungen bzw. die resonanten Phononen aufteilt. Aus $\frac{dN_2}{dt} = 0$ erhält man für den stationären Zustand

$$\frac{N_2}{T_1} = \int_0^\infty \frac{Z(\nu) d\nu}{\tau_{res}(\nu)} \quad (50)$$

Es gilt nach Gl. (44):

$$\tau_{res}^{-1}(\nu) = \frac{4}{D(\nu_0)} g(\nu) (N - 2N_2)$$

Hierbei ist $g(\nu) = \frac{\Gamma^2/4}{\Gamma^2/4 + (\nu - \nu_0)^2}$ mit $\Gamma = \frac{1}{2\pi T_1}$ die Linienformfunktion für die Resonanzstreuung der Phononen. Der Term N in Gl.(44) wurde hier durch $(N - 2N_2)$ ersetzt. Dies berücksichtigt in Gl.(49) auch die stimulierte Emission von Phononen. Damit ergibt sich die Gleichgewichtsbedingung

$$\frac{N_2}{N - 2N_2} D(\nu_0) \Gamma \frac{\pi}{2} = \int_0^\infty Z(\nu) g(\nu) d\nu \quad (51)$$

- b) Wird jetzt die Konzentration N_2 angeregter Zweiniveausysteme, die mit einer Verteilung $Z(\nu)$ resonanter Phononen im Gleichgewicht steht, durch Absorption eines kurzen FIR-Pulses um dN_2 erhöht, ist die Gleichgewichtsbedingung (51) gestört. Es wird eine gewisse Konzentration Δ von resonanten Phononen emittiert, bis die neue Konzentration $N'_2 = N_2 + dN_2 - \Delta$ angeregter Zweiniveausysteme (zusammen mit einer entsprechenden Verteilung $Z'(\nu)$ resonanter Phononen) wieder Bedingung (51) erfüllt. Aus Gründen der Energieerhaltung ist die spektrale Verteilung der erzeugten Phononen die gleiche wie die der absorbierten FIR-Strahlung; bezeichnet man diese mit $g_{FIR}(\nu)$, gilt für die Konzentration $\tilde{Z}(\nu)d\nu$ der erzeugten Phononen im Intervall $[\nu, \nu + d\nu]$: $\tilde{Z}(\nu)d\nu = \frac{\Delta}{\beta} g_{FIR}(\nu) d\nu$. Mit $\beta = \int_0^\infty g_{FIR}(\nu) d\nu$ ist die Normierungsbedingung $\int_0^\infty \tilde{Z}(\nu) d\nu = \Delta$ erfüllt. Die neue Gleichgewichtsbedingung (51) lautet dann:

$$\frac{N'_2}{N - 2N'_2} D(\nu_0) \Gamma \frac{\pi}{2} = \int_0^\infty (Z(\nu) + \tilde{Z}(\nu)) g(\nu) d\nu$$

Sie liefert damit folgende Bedingung für Δ :

$$\frac{N_2 + dN_2 - \Delta}{N - 2(N_2 + dN_2 - \Delta)} D(\nu_0) \Gamma \frac{\pi}{2} = \int_0^{\infty} (Z(\nu) + \frac{\Delta}{\beta} g_{FIR}(\nu)) g(\nu) d\nu \quad (52)$$

- c) Nach den bisherigen Betrachtungen über den stationären Zustand innerhalb des gekoppelten Systems von Zweiniveausystemen und resonanten Phononen wird nun der Fall betrachtet, daß zeitlich veränderliche FIR-Strahlung auf die Probe trifft. Es sei jedoch zunächst bemerkt, daß infolge der Kopplung der Zweiniveausysteme mit dem Schwingungssystem des Kristalls durch Emission und Reabsorption resonanter Phononen die Abhängigkeit der Proben Transmission von der FIR-Pulsenergie wesentlich anders aussieht als im Falle einfacher isolierter Zweiniveausysteme (detaillierte Abhandlungen für den letzteren Fall findet man in [121/]). So kann beispielsweise nie eine volle Transmission erreicht werden; denn dies würde $N_2 = \frac{N}{2}$, d.h. Gleichbesetzung innerhalb der Zweiniveausysteme bedeuten, was jedoch nach Gl.(51) $Z(\nu) \rightarrow \infty$ für $\nu \simeq \nu_0$ implizieren würde. Es müßten also unendlich hohe FIR-Pulsenergien auf den Kristall treffen. (Die Pulsdauer sei als lang gegenüber T_1 betrachtet). Demzufolge hat es auch keinen Sinn, im vorliegenden Fall von einer möglichen Sättigung der FIR-Absorption zu sprechen. Um die Ausbreitung eines FIR-Pulses durch den Kristall zu beschreiben, sei dennoch zunächst der Fall von isolierten Zweiniveausystemen betrachtet. Sie wird in diesem Fall durch folgende Differentialgleichungen festgelegt:

$$\frac{\partial N_2(t, z)}{\partial t} = (N - 2N_2(t, z)) \int_0^{\infty} \sigma_{FIR}(\nu) F(t, z, \nu) d\nu \quad (53a)$$

$$\frac{\partial F(t, z, \nu)}{\partial z} = -\sigma_{FIR}(\nu) F(t, z, \nu) (N - 2N_2(t, z)) \quad (53b)$$

Hierbei bezeichne z die Ausbreitungsrichtung der Strahlung durch den Kristall; $F(t, z, \nu) d\nu$ ist der Photonenfluß innerhalb des Frequenzintervalls $[\nu, \nu + d\nu]$ und $\sigma_{FIR}(\nu)$ gibt den Absorptionsquerschnitt für Strahlung der Frequenz ν an. Es gilt $N \sigma_{FIR}(\nu) = \alpha_{FIR}(\nu)$, wobei $\alpha_{FIR}(\nu)$ der Absorptionskoeffizient für FIR-Strahlung im Fall geringer Pulsenergien ist. Der Term $-2N_2$ berücksichtigt die stimulierte Emission von FIR-Strahlung. Da die beschriebenen Experimente alle bei einer Temperatur von $T = 1.8$ K durchgeführt wurden, gelten die Randbedingungen $N_2(t = 0, z) = 0$ und $Z(t = 0, z, \nu) = 0$.

Die Lösung der Gln.(53) erfolgte numerisch für eine vorgegebene spektrale Photonenflußdichte $F(t, z = 0, \nu)$. Um dabei zusätzlich aber die Kopplung der Zweiniveausysteme mit dem Schwingungssystem des Kristalls zu berücksichtigen, wurde der FIR-Puls zeitlich in einzelne Pulse von der Dauer einer Nanosekunde zerlegt. Jeder dieser

Einzelpulse führt zu einer Erhöhung dN_2 der Konzentration N_2 der angeregten Zweiniveausysteme nach Gl.(53a). Danach läßt sich gemäß Gl.(52) die Konzentration der Phononen berechnen, die emittiert werden müssen, damit ein Gleichgewicht zwischen den Zweiniveausystemen und dem System der resonanten Phononen existiert. (Hier sei daran erinnert, daß der Energieaustausch innerhalb des gekoppelten Systems nach Abschnitt II.2 in Zeiten von weniger als 100 psec erfolgt!) Mit der nach Gl.(52) erhaltenen neuen Konzentration N'_2 angeregter Sauerstoffschwingungen wird das Verfahren für den nächsten Teilpuls nach Gl.(53a) wiederholt. In Verbindung mit Gl.(53b) erhält man schrittweise die spektrale Photonenflußdichte $F(t, z = d, \nu)$ nach dem Kristall der Dicke d . Aus $F(t, z = 0, \nu)$ und $F(t, z = d, \nu)$ läßt sich durch Integration über die Frequenz die (effektive) Transmission als Funktion der Zeit innerhalb des Laserpulses berechnen, zusätzliche Integration über die Zeit liefert die Energietransmission T_{eff} .

Mit dem in a)-c) skizzierten Verfahren wurde die Abhängigkeit der Transmission eines 5 mm dicken Siliziumkristalls mit der Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ von der FIR-Pulsenergie unter den experimentellen Bedingungen aus Abschnitt II.3.1 berechnet. Hierbei wurde auch die dort angesprochene Gaußsche Intensitätsverteilung der FIR-Strahlung innerhalb ihres Strahlquerschnitts mitberücksichtigt. Für die zeitliche Struktur des FIR-Pulses wurde eine idealisierte dreieckige Struktur angenommen, die in Abb. 48 angedeutet ist.

Abb. 49 zeigt das Ergebnis für FIR-Strahlung der Frequenz 29.22 cm^{-1} zusammen mit der experimentell beobachteten energieabhängigen Transmission aus Abb. 46 unter Berücksichtigung, daß die Detektorbandbreite auf 2.8 GHz beschränkt ist. Die Modellrechnung zeigt eine wesentlich stärkere Zunahme der Transmission mit der FIR-Pulsenergie als das experimentelle Ergebnis. Die Annahme, daß keine Dämpfung der resonanten Bandmode innerhalb der Zeitdauer des Laserpulses auftritt, ist somit nicht gerechtfertigt. (Wie bereits erwähnt, wurden die Energien der FIR-Pulse nach Gl.(37) mit dem Korrekturfaktor $\gamma = 3.5$ bestimmt. Nach Abschnitt I.2.2 kann dieser Korrekturfaktor jedoch zwischen 2 und 5 liegen. Höhere Korrekturfaktoren als 3.5 bedeuten aber, daß die experimentell gefundenen Transmissionswerte erst bei höheren Energiewerten gefunden würden als in Abb. 49 angegeben. Eintragen der Modellrechnungsergebnisse in die entsprechend geänderte Skala würde eine noch krassere Abweichung vom Experiment ergeben. Auch bei dem minimalen Korrekturfaktor $\gamma = 2$ liefert die Modellrechnung immer noch eine deutlich stärkere Zunahme der Transmission mit der Pulsenergie als das Experiment.)

Anmerkung 1:

Für die Berechnung der Ausbreitung eines FIR-Pulses durch den Kristall nach Gl.(53) ist entscheidend, daß die resonanten Phononen nicht räumlich das Pumpvolumen verlassen. Dies müßte als ein Energieverlustmechanismus in die Rechnung miteinbezogen werden. Aus Gründen der Schrittweite dz , die für die numerische Lösung von (53b) gewählt wurde, so-

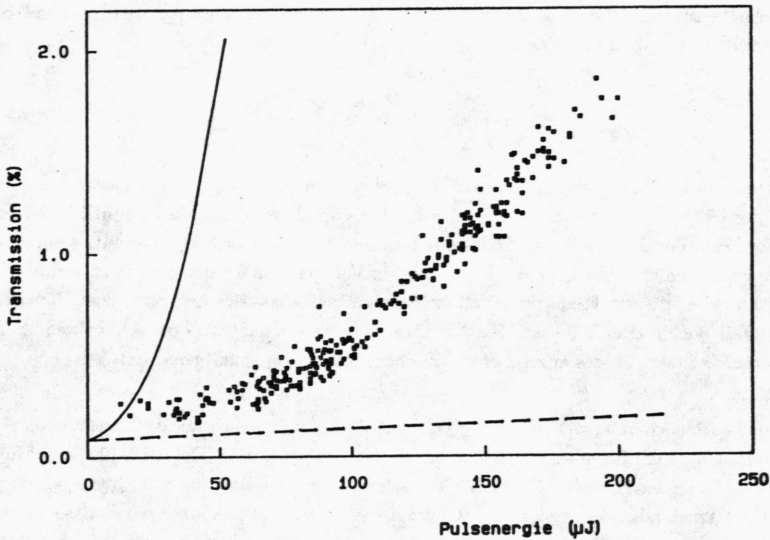


Abb. 49 Transmission einer 5 mm dicken Siliziumprobe mit einer Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ als Funktion der FIR-Pulsenergie für Strahlung der Frequenz 29.22 cm^{-1} . Die durchgezogene Linie ist das Ergebnis der in diesem Kapitel beschriebenen Modellrechnung. Die Bedeutung der gestrichelten Kurve wird in Anmerkung 2 dieses Abschnitts erklärt.

wie der Berücksichtigung der Intensitätsverteilung der FIR-Strahlung über ihren Strahlquerschnitt gilt sogar, daß sie ein Volumen von der Größenordnung einer Kugel mit Radius 0.1 mm nicht verlassen dürfen. Dies läßt sich durch eine kurze Betrachtung der Ausbreitungseigenschaften der resonanten Phononen überprüfen. Infolge der starken Resonanzstreuung breiten sich diese Phononen nur diffusiv im Kristall aus. Analog zur kinetischen Gastheorie [122/ ist die Wahrscheinlichkeit, daß ein Phonon der Frequenz ν bei dreidimensionaler Diffusion die lineare Entfernung x zurückgelegt hat, am größten für die Zeit t_{diff} :

$$t_{diff}(\nu) = \frac{x^2}{6D_{diff}(\nu)} \quad (54)$$

Die Diffusionskonstante $D_{diff}(\nu)$ ist mit der mittleren Schallgeschwindigkeit \bar{v} - für Silizium

errechnet man mit Hilfe der elastischen Konstanten in /99/ den Wert $6.6 \cdot 10^5 \frac{\text{cm}}{\text{sec}}$ - und der Streuzeit $\tau_{\text{res}}(\nu)$ verknüpft durch

$$D_{\text{diff}}(\nu) = \frac{1}{3} \bar{v}^2 \tau_{\text{res}}(\nu) = \frac{\bar{v}^2}{12N} \frac{D(\nu_0)}{g(\nu)} \quad (55)$$

Hierbei wurde Gl.(44) verwendet. Mit der Debyeschen Näherung für die Zustandsdichte $D(\nu_0)$ erhält man - unter Verwendung der mittleren Schallgeschwindigkeit \bar{v} - für Phononen im Zentrum der Resonanz eine Diffusionskonstante $D_{\text{diff}}(\nu_0) = 2.2 \frac{\text{cm}^2}{\text{sec}}$; somit benötigen diese Phononen eine Zeit $t_{\text{diff}} = 7.6 \mu\text{sec}$, um ein Volumen von 0.1 mm Radius zu verlassen. Für Phononen, deren Frequenz 3 Halbwertsbreiten außerhalb des Zentrums liegt, beträgt diese Zeit noch immer 200 nsec. Die Annahme, daß die resonanten Phononen während des Laserpulses innerhalb des angegebenen Volumens bleiben, ist somit gerechtfertigt.

Anmerkung 2:

Es muß auch noch untersucht werden, ob die in Abb. 49 gezeigte nichtlineare Transmission eines 5 mm dicken Siliziumkristalls mit einer Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ für FIR-Strahlung der Frequenz 29.22 cm^{-1} tatsächlich eine Folge eines hohen nichtthermischen Anregungszustandes der resonanten Bandmode ist. Denn man könnte die Frage stellen, ob die Zunahme der Transmission mit zunehmender FIR-Pulsenergie nicht einfach durch eine zunehmende lokale Erwärmung des Kristalls innerhalb des Pumpkanals infolge der dort absorbierten FIR-Energie verursacht werden könnte. Voraussetzung dafür wäre eine lokale Thermalisierung der durch Relaxation der Sauerstoffschwingungen erzeugten Phononen durch Mehrphononenprozesse (siehe Abschnitt II.3.2.2) in Zeiten, die viel kürzer sind als die Dauer des FIR-Pulses. Unter einer lokalen Thermalisierung versteht man eine Umverteilung der Phononenfrequenzen innerhalb eines Teilbereiches des Kristalls zu einer Planckschen Verteilung, die der Temperatur entspricht, die der Kristall in diesem Teilbereich annehmen würde, wenn die gesamte dort absorbierte FIR-Energie in Wärme verwandelt würde. Zu jedem Zeitpunkt während des FIR-Pulses hätte man dann an jedem Ort innerhalb des Pumpkanals eine definierte Temperatur, mit der gemäß der Boltzmannverteilung eine Konzentration N_2 angeregter Zweiniveausysteme verbunden ist. Eine an einem Ort von Null verschiedene Konzentration N_2 angeregter Sauerstoffschwingungen bewirkt aber dort eine Reduktion des Absorptionskoeffizienten $\alpha_{\text{FIR}}(\nu)$ für die FIR-Strahlung gemäß der Beziehung $\alpha_{\text{FIR}}(\nu) = \sigma_{\text{FIR}}(\nu) (N - 2N_2)$ (siehe Gl.(53a)). Eine Erhöhung der FIR-Pulsenergie bedeutet dann eine stärkere lokale Erwärmung des Kristalls und damit auch eine stärkere Reduktion der Absorption; die Transmission des Kristalls nimmt zu.

Eine einfache Abschätzung zeigt jedoch, daß die beobachteten erhöhten Transmissionswerte bei großen FIR-Pulsenergien keine Folge einer solchen lokalen Erwärmung des Kristalls sein können.

Dazu sei angenommen, daß sich die Strahlung in einem Pumpkanal von $1 \times 1 \text{ mm}^2$ Querschnitt durch den Kristall ausbreite. (Dies ist bereits ein wesentlich ungünstigerer Fall, als im Experiment gegeben war; denn dort war die Halbwertsbreite der durch ein Gaußsches Profil beschreibbaren Intensitätsverteilung im Pumpkanal etwa 1.7 mm .) Der Pumpkanal, der ja eine Länge von 5 mm besitzt, sei nun in 10 Stücke der Länge 0.5 mm zerteilt gedacht. Unter der Annahme, daß der Absorptionskoeffizient $\alpha_{FIR}(\nu)$ für die FIR-Strahlung in allen dieser Teilstücke i ($i = 1 \dots 10$) gleich ist und unabhängig von der Pulsenergie der FIR-Strahlung denselben Wert hat, wie er im Falle schwacher Pulsenergien in Abschnitt II.2 ermittelt wurde, nämlich $\alpha_{FIR}(\nu) = 18.5 \text{ cm}^{-1} \cdot g(\nu)$, läßt sich bei einer einfallenden Pulsenergie E (unter Berücksichtigung der spektralen Verteilung der FIR-Strahlung) die in jedem Teilstück i absorbierte Energie E_{abs}^i berechnen. Wenn die gesamte Energie E_{abs}^i in einem solchen Teilstück mit dem Volumen $1 \times 1 \times 0.5 \text{ mm}^3$ vollständig in Wärme verwandelt würde und innerhalb des Teilstücks bliebe, würde dieses die Temperatur

$$T^i = \sqrt[4]{T_o^4 + \frac{4E_{abs}^i}{\delta V}} \quad /123/ \quad (56)$$

erreichen. T_o ist die Temperatur des Kristalls vor Auftreffen des FIR-Pulses ($T_o = 1.8 \text{ K}$), V ist das Volumen des Teilstücks in Bruchteilen eines Mols und δ der Proportionalitätsfaktor, der in dem für $T \ll \Theta_D$ ($\Theta_D \simeq 600 \text{ K}$ ist die Debytemperatur von Silizium /99/) gültigen T^3 -Gesetz der spezifischen Wärme $C_p = \delta T^3$ auftritt. Für Silizium gilt hier $\delta = 0.008 \frac{\text{mJ}}{\text{K}^4 \text{ Mol}} /99/$.

Nach der Boltzmannverteilung läßt sich über T^i die Konzentration N_2^i angeregter Zweiniveausysteme im Teilstück i ermitteln und daraus der reduzierte Absorptionskoeffizient

$$\alpha_{FIR}^i(\nu) = \frac{1 - n^i}{1 + n^i} \alpha_{FIR}(\nu) \quad \text{mit} \quad n^i = e^{-h\nu_o/kT^i} \quad (\text{Boltzmannfaktor}) \quad (57)$$

berechnen. Unter Verwendung dieser reduzierten Absorptionskoeffizienten $\alpha_{FIR}^i(\nu)$ kann man die Transmission des Kristalls für einen hypothetischen Teststrahl geringer Energie, der die gleiche spektrale Verteilung hat wie die FIR-Strahlung, die den Kristall erwärmt, und der zu ihr kollinear durch den Pumpkanal läuft, berechnen. Diese Transmissionswerte sind in Abb. 49 (gestrichelte Kurve) als Funktion der Energie der den Kristall erwärmenden FIR-Pulse aufgetragen. Man erhält z.B. für eine Pulsenergie von $200 \mu\text{J}$ eine Transmission von 0.18% für den hypothetischen Teststrahl. Diese liegt somit um eine Größenordnung unter den gemessenen Werten. Selbst für eine Pulsenergie von $400 \mu\text{J}$ (Pulsenergien dieser Größenordnung würde man nach Gl.(37) erhalten, wenn der maximale Korrekturfaktor $\gamma = 5$ bei der Bestimmung der FIR-Pulsenergie gültig wäre) würde die Transmission des Teststrahls erst 0.25% betragen.

Eine lokale Erwärmung des Kristalls durch die Absorption der FIR-Strahlung kann somit nicht die Ursache der beobachteten Zunahme der Energietransmission mit zunehmender Pulsenergie sein. Es muß vielmehr tatsächlich ein hoher nichtthermischer Anregungszustand der resonanten Bandmode vorhanden sein. Ein direkter Beweis, daß eine schnelle Thermalisierung der resonanten Phononen während des Laserpulses nicht stattfindet, wird außerdem in den beiden folgenden Abschnitten gegeben.

3.2.2 Mögliche Dämpfungsmechanismen für die resonante Bandmode

Eine Dämpfung bzw. ein Zerfall der hochangeregten resonanten Bandmode im System Si:O bedeutet, daß Energie aus dem gekoppelten System von Zweiniveausystemen und resonanten Phononen verlorengeht. Prozesse, die nur die interne Wechselwirkung innerhalb des gekoppelten Systems, nicht aber die darin enthaltene Energie beeinflussen, führen nur zu einer Verbreiterung der Linienform des Streuquerschnitts bzw. der Ferninfrarotabsorption. (So wird beispielsweise die elastische nichtresonante Streuung thermisch vorhandener Phononen an Störstellenschwingungen - sie beeinflussen die Phasenrelaxationszeit T_2 ! - als Ursache der Verbreiterung der FIR-Absorptionslinie von resonanten Bandmoden in dotierten Alkalihalogeniden bei hohen Temperaturen diskutiert /124/). Als Energieverlustmechanismen für die resonante Bandmode kommen in Frage:

- a) Spontaner Zerfall der Sauerstoffschwingung in 2 oder mehr Phononen. Ein solcher Prozeß erscheint möglich, da in der Bindung der Sauerstoffatome an die benachbarten Siliziumatome auch anharmonische Anteile enthalten sind.
- b) Eine Art Differenzprozeß, bei dem ein resonantes Phonon und ein Sauerstoffschwingungsquant vernichtet werden und dabei ein Phonon der doppelten Energie entsteht.
- c) Spontaner Zerfall der resonanten Phononen in 2 niederenergetischere Phononen aufgrund kubisch anharmonischer Terme im Gitterpotential des Siliziumkristalls.
- d) Kombination von 2 resonanten Phononen zu einem Phonon der doppelten Energie (sog. Aufkonversion).

Dies ist der zu c) inverse Prozeß, der bei hohen Phononenbesetzungszahlen bedeutend wird /117/. Als Extremfall wäre sogar eine Thermalisierung der resonanten Phononen durch Mehrphononenprozesse denkbar, wie sie in Anmerkung 2 des vorhergehenden Abschnitts diskutiert wurde.

- e) Inelastische Streuprozesse der resonanten Phononen an im Kristall enthaltenen Defekten, bei denen die Frequenz der Phononen so stark geändert wird, daß sie von den Zweiniveausystemen nicht mehr reabsorbiert werden können.
- f) Inelastische Streuprozesse der resonanten Phononen mit geringer Frequenzverschiebung, so daß die Phononen erst nach vielen Streuprozessen den Resonanzbereich verlassen.

All diesen Prozessen mit Ausnahme von f) ist gemeinsam, daß als Folge Phononen mit Frequenzen weit außerhalb der Resonanzlinie der resonanten Bandmode entstehen. Diese unterliegen dann nicht mehr der Resonanzstreuung und sind somit kein Bestandteil des gekoppelten Systems aus Zweiniveausystemen und resonanten Phononen. Das System hat Energie verloren! Am Ende dieses Abschnitts und im folgenden Kapitel wird nun gezeigt, daß die Prozesse a)-e) als Dämpfungsmechanismen ausgeschlossen werden können, daß aber Prozeß f) einen wesentlichen Zerfallsmechanismus der resonanten Bandmode in Sauerstoff-dotiertem Silizium darstellen könnte.

Berechnungen der Lebensdauer akustischer Phononen in Silizium gegenüber spontanem Zerfall in 2 Phononen aufgrund kubischer Anteile im Gitterpotential wurden von Berke et al. /125/ durchgeführt. Für longitudinale Phononen der Frequenz $\nu_o = 876$ GHz erhält man eine (über alle Ausbreitungsrichtungen gemittelte) Lebensdauer $\tau_L(\nu_o) = 10 \mu\text{sec}$, für schnelle transversale Phononen $\tau_{FT}(\nu_o) = 12.5 \mu\text{sec}$. Liegt ein Gemisch aus longitudinalen und transversalen Phononen dieser Frequenz vor, in dem sich die Phononen untereinander im Gleichgewicht befinden, erhält man eine mittlere Lebensdauer $\bar{\tau}(\nu_o)$, wenn man die Zerfallsraten für die verschiedenen Phononenpolarisationen mit den zugehörigen Zustandsdichten wichtet. In dem hier beschriebenen Fall wird das Gleichgewicht im Gemisch durch die Resonanzstreuung gewährleistet. Es gilt nach /125/

$$\bar{\tau}^{-1}(\nu_o) = \frac{D_L(\nu_o)\tau_L^{-1}(\nu_o) + D_{FT}(\nu_o)\tau_{FT}^{-1}(\nu_o) + D_{ST}(\nu_o)\tau_{ST}^{-1}(\nu_o)}{D_L(\nu_o) + D_{FT}(\nu_o) + D_{ST}(\nu_o)} \quad (58)$$

$D_L(\nu)$, $D_{FT}(\nu)$ und $D_{ST}(\nu)$ sind dabei die Zustandsdichten der longitudinalen und der schnellen bzw. langsamen transversalen Phononen. Vernachlässigt man den Zerfall der langsamen transversalen Phononen $\tau_{ST}^{-1}(\nu_o) \approx 0$, erhält man $\bar{\tau}(\nu_o) \simeq 22 \mu\text{sec}$. Spontaner anharmonischer Zerfall (Prozeß c)) kommt somit als ein Dämpfungsmechanismus für die resonante Bandmode, der bereits während des FIR-Pulses wirksam wird, nicht in Frage.

Bevor nun die anderen Verlustmechanismen betrachtet werden, sei die Ausbreitung von Phononen in Silizium diskutiert, deren Frequenz außerhalb der Resonanzlinie der resonanten Bandmode liegt. Auch diese Phononen breiten sich nicht beliebig weit ballistisch im Kristall aus, sondern unterliegen elastischen nichtresonanten Streuprozessen an Punktdefekten wie Isotopen oder anderen Fremdatomen im Kristall, wenn ihre Frequenz unterhalb der Einsteinoszillationsfrequenz des Defektes liegt. Nach Tamura /126/ beträgt die mittlere Streuzeit $\tau_{el}(\nu)$ für elastische Streuung der Phononen der Frequenz ν an Isotopen in Silizium

$$\tau_{el}^{-1}(\nu) = V_o \frac{4\pi^3}{v^3} \sum_{i=1}^3 f_i \left(1 - \frac{M_i}{M}\right)^2 \nu^4 \quad (59)$$

Dabei ist V_0 das Volumen pro Atom, i indiziert die verschiedenen Isotope; f_i ist der Bruchteil der Isotope mit der Masse M_i und $\bar{M} = \sum_{i=1}^3 f_i M_i$ die mittlere Atommasse. Mit den Isotopenhäufigkeiten von 92.2% Si^{28} , 4.7% Si^{29} und 3.1% Si^{30} erhält man für Phononen der Frequenz 800 GHz eine mittlere Streuzeit von 1.3 μsec und damit eine mittlere freie Weglänge von 8.5 mm. Die relative Konzentration der Sauerstoffatome beträgt $3 \cdot 10^{-3}\%$ und ist somit um den Faktor 3000 kleiner als die der Isotope. Eine Miteinbeziehung in obige Formel wirkt sich somit trotz der größeren Massendifferenz nicht aus. Dies gilt auch für andere im verwendeten Kristall enthaltene Defekte wie Phosphor (10^{14} cm^{-3}), Bor ($< 10^{14} \text{ cm}^{-3}$) und Kohlenstoff ($5 \cdot 10^{16} \text{ cm}^{-3}$).

Bei den Dämpfungsmechanismen a) und e) entstehen Phononen mit Frequenzen unterhalb 800 GHz. Nach der obigen Diskussion besitzen sie somit eine mittlere freie Weglänge von mindestens 8.5 mm. Sie müßten damit den Kristall ballistisch verlassen, wenn ihre Erzeugung 5 mm vom Kristallrand entfernt erfolgt. Nimmt man an, daß die Mechanismen b) oder d) für den Zerfall der resonanten Bandmode verantwortlich sind, würden innerhalb der Zeitdauer des Laserpulses, der eine Halbwertsbreite von typisch 70 nsec besitzt, Phononen der doppelten Frequenz $2\nu_0 = 1.75 \text{ THz}$ erzeugt. Nach Berke et al. [125] beträgt die Lebensdauer dieser Phononen gegenüber anharmonischem Zerfall für longitudinale Phononen $\tau_L(2\nu_0) \simeq 300 \text{ nsec}$ und für schnelle transversale Phononen $\tau_{FT}(2\nu_0) \simeq 400 \text{ nsec}$. Da man aber annehmen kann, daß die Zerfallsprodukte der $2\nu_0$ -Phononen über einen Frequenzbereich verschmiert sind, der groß gegenüber der Halbwertsbreite der Resonanzlinie der resonanten Bandmode ist, entstehen auch hier wieder Phononen mit solchen Frequenzen, daß sie sich ballistisch durch den Kristall ausbreiten können. Dasselbe gilt natürlich erst recht für den Fall, daß während des Laserpulses eine Thermalisierung der durch Relaxation der Sauerstoffschwingungen erzeugten Phononen stattfindet. Durch die Umverteilung der Phononenfrequenzen zu einer Planckschen Verteilung werden wiederum Phononen mit Frequenzen unter 800 GHz erzeugt, die den Kristall ballistisch verlassen können.

Bringt man daher einen Phononendetektor an einer Seite des Kristalls an, der 5 mm von dem Ort entfernt ist, wo die FIR-Strahlung auf den Kristall trifft, müßte dieser Detektor innerhalb einer Zeit von 1 μsec nach Auftreffen des Laserpulses ballistisch den Kristall verlassende Phononen registrieren, wenn einer der Fälle a), b), d) oder e) für die Dämpfung der resonanten Bandmode verantwortlich ist. Das Ergebnis einer experimentellen Überprüfung dieser Fragestellung wird im folgenden Abschnitt geschildert.

3.2.3 Laufzeitexperimente zur Untersuchung der Ausbreitungseigenschaften der durch die FIR-Absorption erzeugten Phononen: Hinweis auf eine räumlich-spektrale Diffusion

Um die Ausbreitung der durch die Relaxation der angeregten Sauerstoffschwingungen erzeugten Phononen zu untersuchen, wurde ein supraleitendes Zinnbolometer auf eine Seitenfläche des Siliziumkristalls aufgebracht. Die experimentelle Konfiguration zeigt Abb. 50.

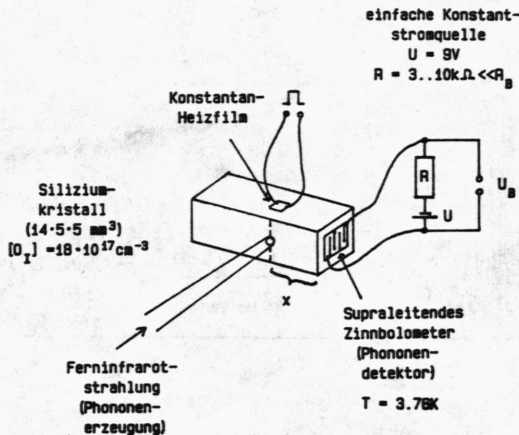


Abb. 50 Experimentelle Anordnung für die Laufzeitexperimente.

Das Bolometer wurde durch Aufdampfen einer $3 \times 4 \text{ mm}^2$ großen Zinnschicht (Dicke 550 \AA) hergestellt, in die zur Erhöhung des Widerstandes R_B (im normalleitenden Zustand) mäanderförmige Strukturen eingeritzt wurden. Der Übergang vom normalleitenden in den supraleitenden Zustand erfolgte innerhalb eines scharfen Temperaturbereichs zwischen $T = 3.74 \text{ K}$ ($R_B = 0$) und $T = 3.84 \text{ K}$ ($R_B = 30 \Omega$). Durch einen speziellen Regelkreis wurde das Bolometer auf einer Temperatur im Übergangsbereich ($T = 3.76 \text{ K}$) gehalten. Auf das Bolometer treffende Phononen bewirken eine Temperaturerhöhung des Bolometers - im metallischen Zinn erfolgt eine schnelle Thermalisierung der Phononen infolge der starken Elektron-Phonon-Kopplung! - und damit eine Widerstandsänderung, die sich bei Verwendung einer Konstantstromquelle als Spannungsänderung registrieren läßt. Die Widerstandsänderung ist in erster Näherung proportional zur Temperaturerhöhung und damit

zu der im Bolometer absorbierten Energie. (Eine detaillierte Abhandlung über supraleitende Bolometer findet sich in [127].) Hat man zusätzlich nur Phononen in einem schmalen Frequenzbereich, ist die momentane Bolometerspannung U_B proportional zu dem momentan auf das Bolometer treffenden Phononenfluß (eine genügend kleine thermische Zeitkonstante des Bolometers sei vorausgesetzt). Über die Zeitabhängigkeit der Spannung U_B läßt sich somit die Zeitabhängigkeit des auf den Detektor treffenden Phononenflusses ermitteln.

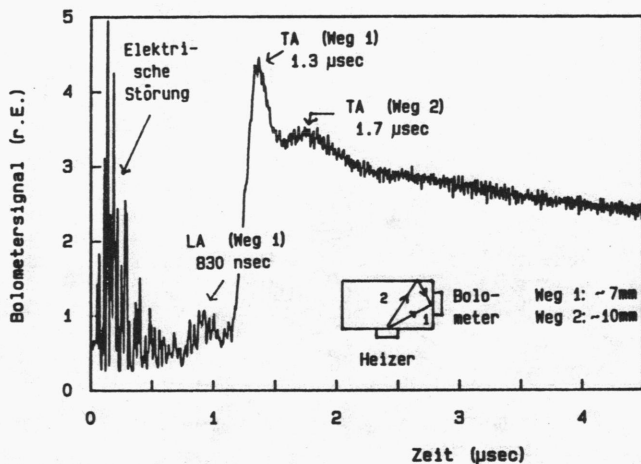


Abb. 51 Laufzeitmessungen mit Wärmepulsphononen (ihre Erzeugung fand bei $t=0$ durch einen 100 nsec langen Heizpuls statt; das Maximum der spektralen Energiedichte lag bei 950 GHz).

Zunächst wurde überprüft, ob mit der Anordnung in Abb. 50 ballistisch laufende Phononen beobachtet werden können. Zu diesem Zweck war zusätzlich ein Konstantanheizfilm (Dicke 100 Å, Fläche $1.5 \times 1.5 \text{ mm}^2$, $R = 24 \text{ } \Omega$) auf eine weitere Seitenfläche des Kristalls aufgedampft. Erhitzt man den Konstantanfilm mit kurzen Strompulslen auf die Temperatur T_H , werden Phononen mit einer der Temperatur T_H entsprechenden Planckschen spektralen Verteilung in den Kristall injiziert [128]. Die Heizertemperatur läßt sich mittels des akustischen Fehlanpassungsmodells [128] abschätzen. Man erhält für die Anordnung in Abb. 50 bei Heizpulsen von 100 nsec Dauer und 20 W Pulsleistung eine Heizertemperatur von $T_H \approx 16 \text{ K}$. Das Maximum der spektralen Energiedichte der in den Kristall injizierten Phononen liegt somit bei 950 GHz. Abb. 51 zeigt das zugehörige Bolometersignal $U_B(t)$

(gemittelt über 50 Pulse). Es sind klar 2 Pulse zu erkennen, die die direkte ballistische Ausbreitung von longitudinalen und transversalen akustischen Phononen vom Heizer zum Detektor anzeigen. Ein dritter Puls ist auf die Ankunft von transversalen Phononen nach einer spiegelnden Reflexion zurückzuführen. Der restliche Teil des Signals wird bestimmt durch die diffusive Ausbreitung von Phononen durch Streuung an Punktdefekten bzw. den Oberflächen des Kristalls.

Abb. 52 a zeigt das Bolometersignal U_B nach Bestrahlung des Siliziumkristalls mit FIR-Pulsen einer Energie von typisch $150 \mu\text{J}$ und einer Frequenz von 29.2 cm^{-1} (gemittelt über 10 Pulse) für einen Abstand $x = 5 \text{ mm}$ zwischen Detektor und Auftreffort der FIR-Strahlung. Die Kurvenform ist typisch für eine diffusive Ausbreitung der Phononen. Messungen bei höherer Zeitauflösung (Abb. 52 b) beweisen eindeutig, daß innerhalb einer Zeit von $5 \mu\text{sec}$ nach dem Laserpuls kein Anzeichen ballistisch laufender Phononen vorhanden ist.

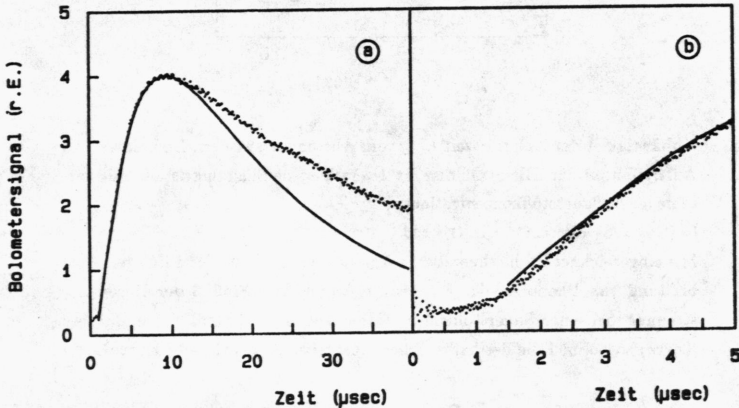


Abb. 52 Bolometersignal nach Bestrahlung des Siliziumkristalls mit FIR-Pulsen (Frequenz 29.2 cm^{-1}) bei $t=0$. Kurve b) zeigt den Anfang des Signals aus a) mit höherer Zeitauflösung. Die durchgezogene Linie ist eine Rechnung nach dem Modell der räumlich-spektralen Diffusion (siehe Text).

In Abb. 53 ist die Ankunftszeit t_{max} des Maximums der Energie, die auf das Bolometer trifft (vgl. Pfeil in Abb. 52 a) als Funktion des Abstandes x zwischen Bolometer und dem Auftreffpunkt der FIR-Strahlung für zwei verschiedene Siliziumkristalle mit verschiedenen Sau-

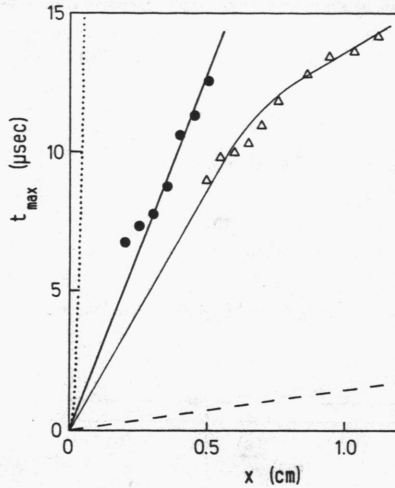


Abb. 53 Abhängigkeit der Ankunftszeit t_{max} vom Abstand x zwischen Bolometer und Auftreffpunkt der FIR-Strahlung für 2 verschiedene Siliziumkristalle mit verschiedenen Sauerstoffkonzentrationen.

(• $[O_I] = 6 \cdot 10^{17} \text{ cm}^{-3}$ $\Delta [O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$)

Mit eingezeichnet ist die theoretische Ankunftszeit im Falle der diffusiven Ausbreitung von Phononen der Frequenz ν_o unter dem Einfluß der Resonanzstreuung bei einer Sauerstoffkonzentration von $18 \cdot 10^{17} \text{ cm}^{-3}$ (gepunktete Kurve) sowie im Falle der ballistischen Ausbreitung (gestrichelte Kurve).

erstoffkonzentrationen aufgetragen. Ebenfalls eingezeichnet ist die theoretische Ankunftszeit t_{max} , die man für den Fall der diffusiven Ausbreitung von Phononen der Frequenz ν_o infolge der resonanten Streuung an den Sauerstoffatomen ($[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$) erwarten würde (gepunktete Kurve). Nach Abschnitt II.3.2.1 ist nämlich die Wahrscheinlichkeit dafür, daß Phononen der Frequenz ν_o im Abstand x von ihrem Erzeugungsort beobachtbar sind, am größten für die Zeit

$$t_{max}(x) = \frac{x^2}{4D_{diff}(\nu_o)} \quad \text{mit} \quad D_{diff}(\nu_o) = 2.2 \frac{\text{cm}^2}{\text{sec}} \quad (60)$$

(Da die experimentelle Anordnung in Abb. 50 eine annähernd zweidimensionale Konfiguration darstellt, steht hier der Faktor 4 im Nenner im Gegensatz zur dreidimensionalen

Betrachtung in Abschnitt II.3.2.1 (Gl.(54)), die den Faktor 6 erfordert /122/. Die errechneten Zeiten übertreffen die gemessenen Zeiten t_{max} zum einen um Größenordnungen, zum anderen erfordern sie eine quadratische Abhängigkeit der Ankunftszeit t_{max} vom Abstand x , während die beobachteten Zeiten zumindest für kleine Abstände x eine lineare Abhängigkeit vom Abstand besitzen (durchgezogene Linien). Die gestrichelte Kurve zeigt als anderen Extremfall das Ergebnis für eine ballistische Ausbreitung der Phononen (berechnet mit der mittleren Schallgeschwindigkeit $\bar{v} = 6.6 \cdot 10^5 \frac{\text{cm}}{\text{sec}}$). Die erhaltenen Zeiten sind hier sehr viel kleiner als die beobachteten.

Die oben genannten Ergebnisse lassen sich aber mit Hilfe des Modells der sog. räumlich-spektralen Diffusion der durch die Relaxation der Sauerstoffschwingungen erzeugten Phononen interpretieren. Dieses Modell wurde erstmals bei Untersuchungen des Resonanzanfangs akustischer 870 GHz Phononen durch das elektronische Zweiniveausystem ($\bar{E}, 2\bar{A}$) optisch angeregter Cr^{3+} -Ionen in Rubin diskutiert /129-131/. Räumlich-spektrale Diffusion bedeutet, daß die Ausbreitung der Phononen durch 2 Einflüsse bestimmt wird, nämlich durch resonante (elastische) Streuung (hier an den Sauerstoffatomen) und zusätzlich durch inelastische Streuprozesse, bei denen sie kleinen Frequenzänderungen unterliegen, die positives oder negatives Vorzeichen haben können. Nimmt man eine mittlere Frequenzverschiebung $\pm \nu_R$ pro inelastischem Streuprozess an und bezeichnet man mit τ_{in} die mittlere Zeit zwischen zwei solchen Streuprozessen (sie sei als unabhängig vom Vorzeichen der Frequenzverschiebung angenommen), läßt sich die zeitliche (spektrale) Veränderung einer gegebenen anfänglichen spektralen Verteilung von Phononen analog einer eindimensionalen räumlichen Diffusion beschreiben. Die räumliche Diffusionskonstante wird hierbei durch die spektrale Diffusionskonstante $D_\nu = \frac{\nu_R^2}{\tau_{in}}$ ersetzt (ν_R entspricht der mittleren freien Weglänge bei der räumlichen Diffusion, τ_{in} der mittleren Zeit zwischen zwei Streuprozessen); sie sei im folgenden als frequenzunabhängig betrachtet. Die spektrale Verteilung der durch die Relaxation der Sauerstoffschwingungen erzeugten resonanten Phononen ist nach Abschnitt II.3.2.1 durch die spektrale Verteilung der absorbierten FIR-Strahlung gegeben ($\simeq 3.6$ GHz!). Ihre Ausbreitung unter dem Einfluß der Resonanzstreuung würde im wesentlichen der gepunkteten Kurve in Abb. 53 folgen. Infolge der spektralen Diffusion verbreitert sich jedoch diese Verteilung. Ein Teil der Phononen wandert (im Frequenzraum) in die Flanken der Resonanzlinie, besitzt damit einen geringeren Streuquerschnitt für die resonante Streuung an den Sauerstoffschwingungen, d.h. größere freie Weglängen und breitet sich somit schneller aus als der Teil der Phononen mit Frequenzen im Resonanzzentrum. Die Ausbreitung der Phononen wird durch die verallgemeinerte Diffusionsgleichung /130/

$$\frac{\partial Z}{\partial t} = D_\nu \frac{\partial^2 Z}{\partial \nu^2} + D_{diff}(\nu) \vec{\nabla}_d^2 Z \quad (61)$$

beschrieben. Hierbei ist $Z d\nu = Z(\nu, \vec{r}, t) d\nu$ wieder die Konzentration der Phononen im Intervall $[\nu, \nu + d\nu]$ am Orte \vec{r} zur Zeit t , $D_{diff}(\nu)$ ist ihre räumliche Diffusionskonstante infolge

der Resonanzstreuung an den Sauerstoffschwingungen nach Gl.(55) und D_ν die erwähnte spektrale Diffusionskonstante; $\vec{\nabla}_d^2$ bezeichne den d-dimensionalen Laplaceoperator. Von Happek et al. /130/ wurde gezeigt, daß im Falle starker spektraler Diffusion, d.h. wenn $D_\nu \gg D_{diff}(\nu_o)$, eine lineare Abhängigkeit der Ankunftszeit t_{max} vom Abstand x zwischen Bolometer und Erzeugungsort der resonanten Phononen existiert. Für den Fall einer zweidimensionalen Geometrie, wie sie annähernd in der Konfiguration in Abb.50 realisiert ist, gilt nach /130/:

$$t_{max}(x) = \frac{0.79 \Gamma}{\sqrt{D_{diff}(\nu_o) D_\nu}} \cdot x \quad (62)$$

Hierbei ist Γ wieder die Breite der Resonanzlinie für die Streuung akustischer Phononen an den Sauerstoffschwingungen. Mit $\Gamma = 4$ GHz und $D_{diff}(\nu_o) = 2.2 \frac{\text{cm}^2}{\text{sec}}$ erhält man für den Fall des mit $18 \cdot 10^{17} \text{ cm}^{-3}$ Sauerstoffatomen dotierten Siliziumkristalls aus der durchgezogenen Linie in Abb. 53 eine spektrale Diffusionskonstante $D_\nu = 1.3 \cdot 10^{28} \frac{\text{Hz}^2}{\text{sec}}$. Daraus Werte für ν_R und τ_{in} zu ermitteln ist, da beide unbekannt sind, nicht möglich. Wäre $\tau_{in} = 10$ nsec, wäre $\nu_R \approx 11.5$ GHz. Die Abweichung der experimentellen Werte vom linearen Zusammenhang zwischen der Ankunftszeit t_{max} und dem Abstand x bei großen Abständen ist eine Frage der endlichen Ausdehnung des Kristalls. Happek et al. /129/ haben gezeigt, daß sich in diesem Fall nach einer genügend langen Zeit t_E eine stationäre räumliche Phononenverteilung im ganzen Kristall ausgebreitet hat, die nur mehr zeitlich exponentiell abnimmt. Es können somit keine längeren Ankunftszeiten t_{max} als t_E gemessen werden!

Ähnliche Resultate, wie sie in Abb.53 bereits für 2 verschiedene Siliziumkristalle gezeigt sind, wurden auch für weitere Kristalle mit noch anderen Sauerstoffkonzentrationen gefunden. Der Wert der spektralen Diffusionskonstante war jedoch für jeden Kristall, die alle von unterschiedlichen Quellen stammten, verschieden. Eine Korrelation mit der Sauerstoffkonzentration konnte dabei nicht gefunden werden /132/. Da die Qualität der verschiedenen Kristalle und die in ihnen enthaltenen Verunreinigungen in den meisten Fällen unbekannt waren, ist auch die mikroskopische Ursache für die inelastischen Streuprozesse zum gegenwärtigen Zeitpunkt noch völlig unbekannt.

Um den Hinweis auf räumlich-spektrale Diffusion anhand der Diskussion von Abb.53 noch zu bekräftigen, wurde in Abb.52 die auf das Bolometer fallende Energie als Funktion der Zeit miteingetragen, die numerisch anhand einer Lösung der Gl.(61) in /130/ für die experimentelle Konfiguration in Abb. 50 für $x = 5$ mm unter Verwendung der obigen spektralen Diffusionskonstante berechnet wurde. (Die endliche Ausdehnung des Bolometers wurde dabei berücksichtigt.) Es zeigt eine gute Übereinstimmung im ersten Teil des Signals. Die Abweichung im zweiten Teil kann eine Folge davon sein, daß der Kristall wie oben erwähnt endlich ist und daß damit insbesondere keine exakte zweidimensionale Geometrie vorhanden ist.

Zusammenfassend gilt: Die Untersuchung der Ausbreitung der durch Relaxation der Sauerstoffschwingungen erzeugten Phononen mittels Laufzeitexperimenten ergab keinen Anhaltspunkt dafür, daß Phononen innerhalb der Zeitdauer von 1 μsec nach dem Laserpuls den Kristall ballistisch verlassen. Die Dämpfungsmechanismen a), b), d) und e) des vorhergehenden Abschnitts können damit gemäß der dortigen Diskussion nicht für den Zerfall der resonanten Bandmode verantwortlich sein. Die Ursache dafür scheinen vielmehr inelastische Streuprozesse zu sein, die kleine Frequenzverschiebungen der resonanten Phononen zur Folge haben, so daß diese allmählich (d.h. nach vielen Streuvorgängen) die Resonanzlinie verlassen und somit langsam dem gekoppelten System von Sauerstoffschwingungen und resonanten Phononen verlorengehen. Welcher mikroskopische Mechanismus für die inelastischen Streuprozesse verantwortlich ist, ist noch nicht bekannt.

3.2.4 Erweiterung der Modellrechnung mit spektraler Diffusion der resonanten Phononen als Dämpfungsmechanismus der resonanten Bandmode

In Abschnitt II.3.2.1 wurde gezeigt, daß die resonanten Phononen, die in Wechselwirkung mit den Zweiniveausystemen stehen, mit denen sich die Sauerstoffschwingungen identifizieren lassen, innerhalb der Dauer des Laserpulses ein Volumen der Größenordnung einer Kugel von 0.1 mm Radius nicht verlassen. Das Näherungsverfahren zur Berechnung der Ausbreitung eines FIR-Pulses durch einen Siliziumkristall nach Gl.(53) erforderte diese Bedingung. Diese Aussage ist aber auch noch im Falle einer räumlich-spektralen Diffusion der resonanten Phononen näherungsweise gültig. Denn nach Abschnitt II.3.2.3 gilt: Werden zu einem Zeitpunkt $t = 0$ an einem Ort resonante Phononen erzeugt, erreicht die Konzentration der Phononen im Abstand x vom Erzeugungsort ihr Maximum zur Zeit $t_{\max} = \frac{0.65 \cdot \Gamma}{\sqrt{D_{\text{diff}}(\nu_0)} D_\nu} \cdot x$. (Der Faktor 0.65 steht anstelle des Faktors 0.79 in Gl.(62), wenn, wie in diesem Fall, eine dreidimensionale Ausbreitung betrachtet wird.) Für $x = 0.1$ mm errechnet man $t_{\max} \simeq 150$ nsec, so daß innerhalb des Laserpulses, dessen komplette Dauer maximal 140 nsec betrug, die Phononen als näherungsweise lokalisiert in obigem Volumen betrachtet werden können. Es genügt also, während des FIR-Pulses nur die spektrale Diffusion der resonanten Phononen zu berücksichtigen. Die Diffusionsgleichung (61) lautet dann:

$$\frac{\partial Z(\nu, t)}{\partial t} = D_\nu \frac{\partial^2 Z(\nu, t)}{\partial \nu^2} \quad (63)$$

Wird zum Zeitpunkt t_0 eine Konzentration Z_0 von Phononen in einem sehr schmalen Frequenzbereich um eine Zentralfrequenz $\tilde{\nu}$ (d.h. im Sinne einer Deltafunktion bei $\tilde{\nu}$) erzeugt, dann ist bekanntlich /122/

$$Z(\nu, t) = \frac{Z_0}{2\sqrt{\pi D_\nu(t - t_0)}} \exp\left(-\frac{(\nu - \tilde{\nu})^2}{4 D_\nu(t - t_0)}\right) \quad \forall t > t_0 \quad (64)$$

die Lösung der Diffusionsgleichung (63). Insbesondere ist die Normierungsbedingung $\int_0^\infty Z(\nu, t) d\nu = Z_0 \forall t > t_0$ erfüllt. Für die Halbwertsbreite $\Delta\nu_p$ dieser spektralen Phononenverteilung (64) gilt: $\Delta\nu_p = 4\sqrt{\ln 2 \sqrt{D_\nu(t-t_0)}} \quad \forall t > t_0$; d.h. 100 nsec nach der (monochromatischen) Erzeugung der Phononen beträgt die Breite ihrer spektralen Verteilung bereits $\Delta\nu_p \simeq 120$ GHz (bei $D_\nu = 1.3 \cdot 10^{28} \frac{\text{Hz}^2}{\text{sec}}$), was sehr groß gegenüber der spektralen Breite $\Delta\nu_L \simeq 3.6$ GHz der FIR-Strahlung ist. Die Anfangsverteilung der durch Relaxation der Sauerstoffschwingungen erzeugten Phononen, die ja nach Abschnitt II.3.2.1 gegeben ist durch die spektrale Verteilung der absorbierten Laserstrahlung, kann somit als näherungsweise delpulsförmig betrachtet werden, ihre zeitliche Entwicklung ist dann durch Gl.(64) gegeben.

Nach Abschnitt II.3.2.1 wurde zur Berechnung der Ausbreitung eines FIR-Pulses durch den Kristall nach Gl.(53) der Puls zeitlich in einzelne Teilpulse von der Dauer einer Nanosekunde zerlegt. Der zu einem Zeitpunkt $t_i = i$ nsec einfallende Teilpuls führt zu einer Erhöhung dN_2 der Konzentration N_2 angeregter Zweiniveausysteme nach Gl.(53a). Über Gl.(52) erhält man die Konzentration Δ_i der Phononen, die zum Zeitpunkt t_i emittiert werden müssen, damit die Gleichgewichtsbedingung (51) erfüllt ist. Im Gegensatz zum Abschnitt II.3.2.1, wo man auf diese Art und Weise zu einem Zeitpunkt t_j folgende spektrale Verteilung der Phononenkonzentration

$$Z(\nu, t_j) = \sum_{i=1}^{j-1} \frac{\Delta_i}{\beta} g_{\text{FIR}}(\nu) \quad (65)$$

vorliegen hat, die ja nach Gl.(52) zusammen mit der Erhöhung dN_2 der Konzentration angeregter Zweiniveausysteme durch den FIR-Teilpuls zur Zeit t_j die Konzentration Δ_j emittierter Phononen zu diesem Zeitpunkt bestimmt, muß jetzt die spektrale Diffusion der Phononen nach Gl.(64) berücksichtigt werden. Es gilt somit:

$$Z(\nu, t_j) = \sum_{i=1}^{j-1} \frac{\Delta_i}{2\sqrt{\pi D_\nu(t_j - t_i)}} \exp\left(-\frac{(\nu - \nu_L)^2}{4D_\nu(t_j - t_i)}\right) \quad (66)$$

Ansonsten ist die Durchführung der Berechnung der Ausbreitung eines FIR-Pulses durch den Kristall völlig identisch zu der, die in Abschnitt II.3.2.1 für den Fall verschwindender Dämpfung der resonanten Bandmode beschrieben wurde.

Für einen 5 mm dicken Siliziumkristall mit der Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$, an dem ja die Experimente durchgeführt wurden, wurde die Abhängigkeit seiner Transmission für FIR-Strahlung der Frequenz 29.22 cm^{-1} von der Pulsenergie unter den oben genannten Nebenbedingungen und den experimentellen Bedingungen aus Abschnitt II.3.1 berechnet. Für die zeitliche Struktur des FIR-Pulses wurde wieder die in Abb.48 a skizzierte dreieckige Struktur angenommen; die annähernd Gaußsche Intensitätsverteilung der

FIR-Strahlung innerhalb ihres Strahlquerschnitts wurde berücksichtigt. Nun war gemäß der Diskussion in den Abschnitten I.3.1 und II.3.1 der exakte Wert der FIR-Pulsenergie nicht bekannt. Ihre Berechnung erfolgte nach Gl.(37), die einen Korrekturfaktor γ enthält, der im Bereich zwischen 2 und 5 liegt. Dieser Wert wurde in der Modellrechnung als freier Parameter betrachtet. Für $\gamma = 3.5$ erhält man eine sehr gute Übereinstimmung zwischen der experimentell beobachteten Abhängigkeit der Transmission von der Pulsenergie und dem Ergebnis der Modellrechnung. Dies ist in Abb. 46 dargestellt. (Daß der Detektor nur Strahlung einer Bandbreite von 2.8 GHz registrieren konnte, wurde in der Modellrechnung berücksichtigt.) Die Abweichungen zwischen der experimentell beobachteten und der berechneten Transmission bei kleinen Pulsenergien können eine Folge davon sein, daß in diesem Fall die zeitliche Struktur des Laserpulses nicht mehr die idealisierte dreieckige Form, wie sie in Abb. 48 a angedeutet ist, besaß, sondern daß sie (siehe Abschnitt I.2.3.2) vielmehr stark strukturiert war und aus einigen wenigen Einzelpulsen bestand, deren Intensität natürlich stärker war als die, die für eine dreieckige Struktur aus der Pulsenergie berechnet wurde.

Eine befriedigende Übereinstimmung zwischen Experiment und Modellrechnung erhält man auch für die in Abb. 47 gezeigten Fälle, bei denen die energieabhängige Transmission für verschiedene Laserfrequenzen in der Nähe des Zentrums der resonanten Bandmode bestimmt wurde; wie in Abschnitt II.3.1 bereits diskutiert, war in diesem Fall keine Einengung der Detektorbandbreite durch ein Fabry-Perot-Interferometer vorhanden. Das zusätzliche Auftreten einer zweiten Komponente mit einer Frequenz von 28.5 cm^{-1} im Emissionsspektrum des CH_3F -Raman-Lasers bei Pulsenergien über $125 \mu\text{J}$, die nicht im Kristall absorbiert wurde, führt daher zu einer Abweichung zwischen experimentell beobachteter und berechneter Transmission bei hohen Pulsenergien. Es sei bemerkt, daß neben der Bandbreite des Detektors die Laserfrequenz ν_L die einzige Größe war, die bei der Berechnung der durchgezogenen Kurven in Abb. 47 gegenüber der Rechnung im Falle von Abb. 46 geändert wurde.

Die zeitliche Änderung der Transmission innerhalb eines FIR-Pulses mit der Frequenz $\nu_L = 29.22 \text{ cm}^{-1}$ und einer Pulsenergie von $110 \mu\text{J}$, wie sie in Abb. 48 c gezeigt ist, soll hier noch kurz genauer betrachtet werden. In Abb. 64 ist die gemessene Transmission für diesen Fall noch einmal aufgezeichnet (durchgezogene Linie). Die gestrichelte Linie ist das Ergebnis der Modellrechnung, wie sie am Anfang dieses Kapitels beschrieben wurde. Auch sie reproduziert sehr gut die gemessene zeitliche Änderung der Transmission. (Die starken Oszillationen der Transmission gegen Ende des FIR-Pulses resultieren aus elektrischen Schwingungen des Verstärkers, der bei der Messung der transmittierten Pulsintensität verwendet wurde.)

Die gepunktete Kurve zeigt das Ergebnis der Modellrechnung für den Fall, daß der Mechanismus e) aus Abschnitt II.3.2.2 für den Zerfall der resonanten Bandmode verantwortlich ist, d.h. daß inelastische Streuprozesse mit einer Rate τ_{in} so starke Frequenzänderungen der durch die Relaxation der Sauerstoffschwingungen erzeugten Phononen bewirken, daß

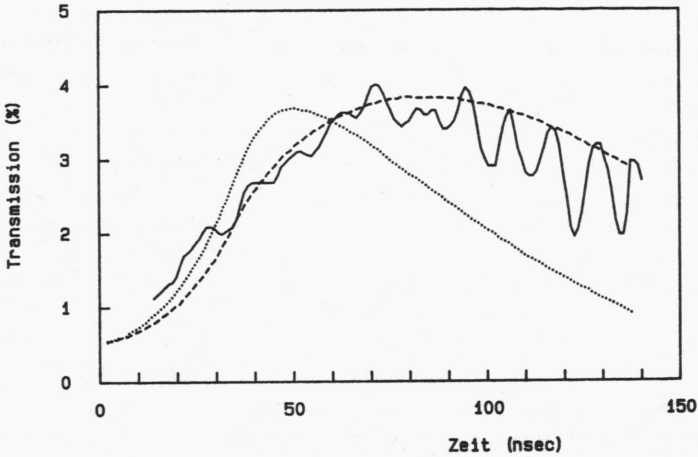


Abb. 54 Zeitliche Änderung der Transmission eines 5 mm dicken Siliziumkristalls mit einer Sauerstoffkonzentration $[O_I] = 18 \cdot 10^{17} \text{ cm}^{-3}$ für einen FIR-Puls einer Frequenz von 29.22 cm^{-1} und einer Energie von $110 \mu\text{J}$. Die gestrichelte bzw. gepunktete Kurve ist das Ergebnis der Modellrechnung bei Berücksichtigung der spektralen Diffusion bzw. des Dämpfungsmechanismus e) aus Abschnitt II.3.2.2.

diese nicht mehr der Resonanzstreuung unterliegen. In diesem Fall gilt für die spektrale Verteilung der Phononen, die innerhalb der Resonanzlinie der resonanten Bandmode zum Zeitpunkt t_j übriggeblieben sind und die durch ihre Wechselwirkung mit den Zweiniveausystemen die Konzentration der angeregten Zustände dieser Systeme nach Gl.(51) und damit die Transmission für die FIR-Strahlung bestimmen:

$$Z(\nu, t_j) = \sum_{i=1}^{j-1} \frac{\Delta_i}{\beta} g_{FIR}(\nu) \exp\left(-\frac{t_j - t_i}{\tau_{in}}\right) \quad (67)$$

Mit $\tau_{in} \approx 7 \text{ nsec}$ reproduziert die berechnete Kurve auch in diesem Fall gut die experimentell beobachtete energieabhängige Transmission aus Abb. 46. Die zeitlich aufgelöste Transmission (gepunktete Kurve in Abb. 64) zeigt jedoch einen wesentlich stärkeren Abfall im zweiten Teil des Laserpulses als experimentell beobachtet wurde. Ähnliche Ergebnisse sind auch für die restlichen Dämpfungsmechanismen aus Abschnitt II.3.2.2 zu erwarten. Dieses Ergebnis liefert (neben den Laufzeitexperimenten aus Abschnitt II.3.2.3) einen weiteren Hinweis, daß

die spektrale Diffusion der resonanten Phononen der hauptsächlich für den Zerfall der resonanten Bandmode verantwortliche Prozeß ist. Denn bei der spektralen Diffusion verlangsamt sich (wie bei jeder Diffusion) die Verlustrate einer gegebenen Anfangsverteilung resonanter Phononen aus einem vorgegebenen Frequenzintervall mit zunehmender Zeit, während sie beim oben betrachteten Prozeß ja konstant bleibt. Dies führt zu einem langsameren Abfall der Transmission im zweiten Teil des Laserpulses im Fall der spektralen Diffusion, als es für einen Dämpfungsprozeß mit konstanter Verlustrate der Fall wäre.

3.3 Ausblick

In diesem Kapitel wurde erstmals eine nichtlineare Absorption von FIR-Strahlung durch eine resonante Bandmode in Kristallen mit Punktdefekten demonstriert. Modellsystem war ein mit Sauerstoff dotierter Siliziumkristall, in dem der in interstitieller Form vorliegende Sauerstoff zum Auftreten einer resonanten Bandmode bei 29.2 cm^{-1} führt. Mit Hilfe des im Rahmen dieser Arbeit entwickelten abstimmbaren CH_3F -Raman-Lasers konnte eine Erhöhung der Transmission eines Sauerstoff-dotierten Siliziumkristalls für FIR-Strahlung obiger Frequenz bei hohen Pulsenergien beobachtet werden. Es wurde außerdem eine Modellrechnung vorgestellt, die die nichtlineare Absorption von FIR-Strahlung durch ein System von homogen in einem Kristall verteilten Zweiniveausystemen beschreibt, das sich durch Emission und Reabsorption resonanter Phononen im (partiellen) Gleichgewicht mit dem Schwingungssystem des Kristalls befindet. (Die resonante Bandmode in Sauerstoff-dotiertem Silizium läßt sich als ein solches System interpretieren!) Vergleich von Modellrechnung und experimentell beobachteter Transmission gab Anlaß zu dem Schluß, daß eine Dämpfung der resonanten Bandmode bereits während des etwa 140 nsec dauernden Laserpulses eintritt. Von allen betrachteten Dämpfungsmechanismen erwies sich einzig die Vorstellung einer spektralen Diffusion der durch die Relaxation der Sauerstoffschwingungen erzeugten Phononen als verträglich mit den experimentellen Ergebnissen, insbesondere mit Experimenten zur Ausbreitung dieser Phononen, so daß spektrale Diffusion der resonanten Phononen als Ursache des Zerfalls der resonanten Bandmode vorgeschlagen wurde.

Einen direkten Beweis für die spektrale Diffusion der resonanten Phononen stellen die genannten Experimente jedoch nicht dar. Um sie direkt nachzuweisen, sei folgendes Experiment vorgeschlagen:

Neutrale Donatoren bzw. Akzeptoren in Halbleitern sind in der Lage, weitere freie Elektronen bzw. Löcher an sich zu binden und sog. D^- - bzw. A^+ -Zentren [133] zu bilden. Die Bindungsenergie für diese zusätzlichen Ladungsträger lassen sich in der effektiven Massennäherung analog zum Fall eines H^- -Ions näherungsweise berechnen. Man erhält für A^+ -Zentren in Silizium eine Bindungsenergie von etwa 2 meV; Berücksichtigung von

'central-cell'-Effekten führt zu einer Erhöhung dieses Wertes je nach Typ der Störstellen /134/. Da bei tiefen Temperaturen ($T \leq 1.8$ K) jedoch keine freien Ladungsträger vorhanden sind, müssen diese, damit sich D^- - oder A^+ -Zentren bilden können, beispielsweise durch Einstrahlen von Licht mit einer Energie größer als die Bandlücke des Halbleiters bzw. von Infrarotstrahlung mit einer Energie größer als die Bindungsenergie neutraler Störstellen erzeugt werden. Man erhält dann eine (von der eingestrahlten Lichtintensität und der Störstellenkonzentration abhängige) Gleichgewichtskonzentration von D^- - bzw. A^+ -Zentren. Treffen Phononen mit einer Energie größer als die Bindungsenergie der zusätzlichen Ladungsträger an den D^- bzw. A^+ -Zentren auf solche Zentren, werden diese neutralisiert und zusätzliche freie Ladungsträger erzeugt. Man beobachtet eine Erhöhung der Leitfähigkeit. Sie wird als phononeninduzierte Leitfähigkeit (PIC) bezeichnet. Burger /134/ hat eine detaillierte Analyse der Bindungsenergie der zusätzlichen Ladungsträger an den D^- - bzw. A^+ -Zentren nach der PIC-Methode durchgeführt. (Die Phononen wurden von ihm quasimonochromatisch mittels supraleitender Tunneldioden erzeugt.) Er fand für mit dem Akzeptor Bor in einer Konzentration von 10^{16} cm^{-3} dotierte Siliziumkristalle eine Bindungsenergie der zusätzlichen Löcher an den Bor^+ -Zentren von 3.85 meV bzw. 930 GHz. Würde man nun Sauerstoff-dotierte Siliziumkristalle zusätzlich mit Boratomen in der angegebenen Konzentration dotieren, könnte man durch Beobachtung des PIC-Signals während und nach Auftreffen des FIR-Pulses auf den Kristall die Information darüber erhalten, ob und wann Phononen einer Frequenz größer als 930 GHz im Kristall auftreten. (Die FIR-Strahlung mit einer Frequenz von 876 GHz bewirkt keine Neutralisation der Bor^+ -Zentren.) Im Falle eines positiven Ergebnisses hätte man einen direkten Nachweis einer Frequenzänderung von 54 GHz der bei 876 GHz erzeugten Phononen.

Bei Verwendung des Akzeptors Indium statt Bor hätte man auch die Möglichkeit, das Auftreten von Aufkonversionsprozessen zu überprüfen. Denn die Indium $^+$ -Zentren besitzen eine Bindungsenergie für das zusätzliche Loch von etwa 5.9 meV bzw. 1.4 THz /134/. Die diskutierte spektrale Diffusion, die ja nach Abschnitt II.3.2.3 in 100 nsec nur eine Verbreiterung der bei $\nu_0 = 876$ GHz erzeugten schmalbandigen Phononenverteilung auf 120 GHz bewirkt, kann ein PIC-Signal während des Laserpulses nicht erzeugen. Treten jedoch Aufkonversionsprozesse bzw. Thermalisierungsprozesse bereits während des Laserpulses auf, erhält man zu diesem Zeitpunkt auch ein PIC-Signal, da bei diesen Prozessen ja Phononen der Frequenz $2\nu_0 = 1.75$ THz entstehen.

Weiterhin wäre die bislang unbekannte mikroskopische Ursache der spektralen Diffusion der resonanten Phononen zu klären. So könnte beispielsweise mittels gezielter Störstellenspektroskopie (z.B. durch photothermale Infrarotspektroskopie /135/) Art und Konzentration der in den Kristallen enthaltenen Störstellen bestimmt und eine mögliche Korrelation mit der spektralen Diffusionskonstante ermittelt werden.

Von Interesse wäre es auch, die in diesem Kapitel vorgestellten Experimente mit FIR-

Strahlung der gleichen Pulsenergie, aber wesentlich geringerer spektraler Breite $\Delta\nu_L$ zu wiederholen. In diesem Fall würden sehr viel größere Phononenbesetzungszahlen $p(\nu_o)$ als in den vorliegenden Experimenten erreicht. (Die Phononenbesetzungszahl $p(\nu_o)$ ist definiert durch $p(\nu_o) = \frac{Z(\nu_o)}{D(\nu_o)\Delta\nu_L} / 117$, wobei $Z(\nu_o)$, wie gehabt, die Konzentration der durch die Relaxation der Sauerstoffschwingungen erzeugten Phononen ist und $D(\nu_o)$ ihre Zustandsdichte darstellt.) Aufkonversionsprozesse der resonanten Phononen wären dann sehr viel wahrscheinlicher $/117/$. Es wäre zu untersuchen, ob und in welcher Weise Änderungen gegenüber den in dieser Arbeit diskutierten Ergebnissen auftreten. (Man vergleiche dazu die Diskussion, ob eine lokale Erwärmung des Siliziumkristalls durch die Absorption der FIR-Strahlung nicht die Ursache der Erhöhung der Transmission mit zunehmender FIR-Pulsenergie sein könnte, in Anmerkung 2 von Abschnitt II.3.2.1.)

Letztendlich wäre es interessant, die in diesem Kapitel beschriebenen Absorptionsexperimente auch an resonanten Bandmoden in anderen Kristallsystemen durchzuführen. Beispiele wären die Systeme $\text{MgF}_2:\text{Eu}^+$ (charakteristische Daten der resonanten Bandmode: $\nu = 15.95 \text{ cm}^{-1}$, $\Gamma \leq 0.15 \text{ cm}^{-1}$), $\text{NaCl}:\text{Cu}^+$ ($\nu = 23.57 \text{ cm}^{-1}$, $\Gamma = 0.4 \text{ cm}^{-1}$), $\text{KBr}:\text{Li}$ ($\nu = 16.07 \text{ cm}^{-1}$, $\Gamma = 0.43 \text{ cm}^{-1}$) oder $\text{KJ}:\text{Ag}^+$ ($\nu = 17.23 \text{ cm}^{-1}$, $\Gamma = 0.45 \text{ cm}^{-1}$) $/111/$. Die resonanten Bandmoden in all diesen Systemen liegen in einem Frequenzbereich, der mit Hilfe eines $^{13}\text{CH}_3\text{F}$ -Raman-Lasers (siehe Abschnitt I.2.1.2) zugänglich wäre.

Auch elektronische Zweiniveausysteme in Kristallen, die durch Emission und Reabsorption resonanter Phononen an das System der Gitterschwingungen gekoppelt sind, könnten mit den in diesem Kapitel beschriebenen Experimenten untersucht werden. Ein Beispiel wäre das Zweiniveausystem (\bar{E} , $2\bar{A}$) optisch angeregter Cr^{3+} -Ionen in Rubin, das einen Energieabstand von 29.25 cm^{-1} besitzt. Die Linienbreite der FIR-Absorptionslinie beträgt $0.35 \text{ GHz} /136/$.

Eine andere Anwendungsmöglichkeit für das hier vorgestellte FIR-Lasersystem wäre die Untersuchung von Ionenleitern wie RbAg_4I_5 oder ferroelektrischen Kristallen wie TGS und TGSe. Diese Substanzen zeigen eine Reihe scharfer Absorptionslinien im fernen Infrarot $/137,137,139/$. Sättigungsexperimente könnten Aufschluß geben, ob das Potential der die Absorption verursachenden Oszillatoren einen anharmonischen Charakter hat.

Auch der Hochdruck- CO_2 -Laser allein bietet Einsatzmöglichkeiten in der nichtlinearen Spektroskopie. Mit seiner Hilfe könnte beispielsweise das Sättigungsverhalten infrarotaktiver lokalisierter Moden in gestörten Kristallen untersucht werden. Ein Beispiel dafür ist das System $\text{CaF}_2:\text{H}^-$, das eine schmalbandige lokalisierte Mode bei 965.1 cm^{-1} besitzt $/140/$.

Zusammenfassung

Die vorliegende Arbeit beschreibt die Entwicklung eines abstimmbaren Hochleistungslasersystems für den ferninfraroten Spektralbereich. Das Lasersystem besteht aus einem Hochdruck-CO₂-Laser und einem CH₃F-Raman-Laser und ermöglicht eine abstimmbare Erzeugung von FIR-Strahlung in einzelnen Intervallen im Frequenzbereich von 23 cm⁻¹ bis 45 cm⁻¹. Dabei besitzen die Abstimmintervalle eine Breite von 1.2 cm⁻¹ und untereinander einen Abstand von 0.4 cm⁻¹. Mit Hilfe des Lasersystems wurde erstmals die nichtlineare Absorption von FIR-Strahlung bei starker Anregung einer resonanten Bandmode in einem gestörten Kristall, deren Frequenz im Bereich der hochfrequenten akustischen Phononen des ungestörten Kristalls liegt, beobachtet.

Das der Erzeugung der FIR-Strahlung zugrundeliegende Prinzip war die stimulierte Ramanstreuung der Strahlung eines CO₂-Lasers in CH₃F-Gas. Die FIR-Strahlung kann dabei abstimmbar erzeugt werden, wenn der CO₂-Laser abstimmbar ist. Es wurde daher im Rahmen dieser Arbeit zunächst ein Hochdruck-CO₂-Laser mit einem Betriebsdruck von 20 atm entwickelt. Bei diesem Druck sind die Rotations-Schwingungs-Übergänge eines CO₂-Moleküls so stark verbreitert, daß ihr Überlapp eine kontinuierliche Emissionscharakteristik des Lasers zur Folge hat. Der Laser wurde durch eine selbständige Gasentladung mit UV-Vorionisierung angeregt. Seine Konstruktion enthielt dabei zwei wesentliche Neuerungen gegenüber den bisher veröffentlichten Konstruktionen solcher Laser:

- Die zur UV-Vorionisierung verwendeten Funkenstrecken waren parallel geschaltet. Die dadurch erreichte Verringerung der Induktivität des Vorionisierungsschaltkreises gegenüber der bisher benutzten seriellen Schaltung der Funkenstrecken führt zu sehr kurzen und intensiven UV-Strahlungspulsen sowie zu einer synchronen Vorionisierung des gesamten Entladungsvolumens. Eine direkte Kopplung von Hauptentladungs- und Vorionisierungskreis war somit möglich. Sie wurde mittels Kondensatoren innerhalb der Hochdruckkammer durchgeführt und erlaubte daher eine sehr kompakte Laserkonstruktion, die auch nur eine Hochspannungsversorgung nötig machte.
- Die Verwendung von 2 Entladungsstrecken ergab eine deutliche Ausweitung des Abstimmbereichs und eine Erhöhung der Ausgangspulsenergie bei gleichzeitiger Reduktion der Versorgungsspannung gegenüber dem Betrieb mit nur einer Entladungsstrecke, wie er bisher üblich war.

Der Hochdruck-CO₂-Laser ermöglichte eine kontinuierliche Abstimmung in allen Emissionsbereichen eines CO₂-Lasers. Insbesondere war die Emissionsfrequenz des Lasers eindeutig

durch die Winkelposition des im Resonator zur Frequenzselektion verwendeten Reflexionsgitters bestimmt. 'Frequency pulling', das häufig bei den bisher üblichen Hochdruck-CO₂-Lasern beobachtet worden war, die einen Betriebsdruck von 10-12 atm besaßen, trat nicht auf. Der Hochdruck-CO₂-Laser wurde dabei mit einem Gemisch aus CO₂, N₂ und He in der Zusammensetzung CO₂ : N₂ : He = 3:1:60 bei einem Druck von 20 atm betrieben. Die Hochspannungsversorgung erfolgte durch einen zweistufigen Marxgenerator mit einer Kapazität von 10 nF pro Stufe. Typische Ladespannungen lagen bei 50-60 kV. Das durch die Elektroden mit Changprofil gebildete aktive Volumen war $1 \times 1 \times 80 \text{ cm}^3$.

Die Eigenschaften des Lasers im einzelnen waren: Der Abstimmbereich betrug 16 cm^{-1} im 9R-Zweig (1069 cm^{-1} - 1085 cm^{-1}) und 15 cm^{-1} im 10R-Zweig (966 cm^{-1} - 981 cm^{-1}) bei einer Entladungsenergiedichte von $21 \frac{\text{J}}{\text{atm}}$, bzw. 16 cm^{-1} im 9P-Zweig (1040 cm^{-1} - 1056 cm^{-1}) und 18 cm^{-1} im 10P-Zweig (936 cm^{-1} - 954 cm^{-1}) bei einer Entladungsenergiedichte von $26 \frac{\text{J}}{\text{atm}}$. Die Emissionscharakteristik war in den R-Zweigen völlig glatt, während in den P-Zweigen noch eine geringfügige Modulation (Tiefe etwa 15%) infolge der dort größeren Abstände der Rotations-Schwingungs-Übergänge vorhanden war. Eine Frequenzvariation war dabei in Schritten von 0.04 cm^{-1} möglich. Die spektrale Breite der emittierten Strahlung (FWHM) betrug in den Bereichen hoher Verstärkung 0.16 cm^{-1} . Eine obere Grenze für die Pulsenergie (300 mJ) war nur durch die Zerstörungsschwelle der NaCl-Brewsterfenster, die den optischen Zugang zur Hochdruckkammer ermöglichten, festgelegt. Die Dauer (FWHM) der Laserpulse betrug 130 nsec bei einem Betrieb weit über der Schwelle, d.h. bei hohen Kleinsignalverstärkungen im Zentrum der Emissionsbereiche; die Pulsform war dabei nahezu strukturlös. Bei einem Betrieb in der Nähe der Schwelle, d.h. am Rande der Abstimmbereiche, wiesen die Pulse dagegen bei einer Dauer von 190 nsec eine stark strukturierte Form auf, da das Lasermedium selbst eine Phasenkopplung der wenigen longitudinalen Moden erzeugt, die innerhalb der (bei einem Betrieb nahe der Schwelle reduzierten) Bandbreite enthalten sind. Ein spektral breitbandiger Betrieb war durch Ersetzen des Reflexionsgitters durch einen hochreflektierenden Spiegel möglich. Die Emission erfolgte dann nahezu vollständig im Zentrum des 10R-Zweiges bei einer spektralen Breite (FWHM) von etwa 2 cm^{-1} .

Der CH₃F-Raman-Laser wurde mit der Strahlung des 9R-Emissionszweiges des Hochdruck-CO₂-Lasers gepumpt. Die Kleinsignalverstärkung für die stimulierte Ramanstreuung der CO₂-Laserstrahlung wird durch einen doppelten quasiresonanten Ramanprozeß bestimmt: der Ramanübergang eines CH₃F-Moleküls kann nämlich entweder bei gleichzeitiger Emission eines FIR-Photons durch Rotationsübergänge im Schwingungsgrundzustand oder durch entsprechende Übergänge im ersten angeregten Schwingungsgrundzustand erfolgen. Die Interferenz der Wahrscheinlichkeitsamplituden der beiden Prozesse hat wesentlichen Einfluß auf die Abstimmcharakteristik des CH₃F-Raman-Lasers. So war eine abstimbare Erzeugung von FIR-Strahlung im Frequenzbereich von 23 cm^{-1} bis 45 cm^{-1} in einzelnen Intervallen möglich, die eine Breite von etwa 1.2 cm^{-1} und untereinander einen Abstand von 0.4 cm^{-1}

besaßen. Innerhalb der Intervalle bestand ein linearer Zusammenhang zwischen der Frequenz des CO_2 -Lasers und der Frequenz der FIR-Strahlung. Eine Frequenzvariation der FIR-Strahlung konnte dabei in Schritten von 0.04 cm^{-1} entsprechend der minimal möglichen Frequenzänderung der CO_2 -Laserstrahlung erfolgen. Die Bandbreite der emittierten Strahlung war $(4 \pm 0.5) \text{ GHz}$. Die Dauer (FWHM) der FIR-Pulse betrug typischerweise 70 nsec und war somit deutlich geringer als die Dauer der CO_2 -Laserpulse. Ursache ist die Reduktion der Kleinsignalverstärkung während des CO_2 -Laserpulses infolge einer Entvölkerung des Schwingungsgrundzustandes durch die stimulierte Ramanstreuung selbst. Innerhalb der Abstimmintervalle wurde eine starke Variation der Pulsenergie der FIR-Strahlung mit ihrer Frequenz gefunden. So wurden im Bereich großer Kleinsignalverstärkungen FIR-Pulsenergien von über $500 \mu\text{J}$ erreicht. Der Wirkungsgrad für die Umsetzung von CO_2 -Laser-Photonen in FIR-Photonen lag dabei im Bereich von $4\% - 11\%$. Eine starke Abnahme der FIR-Pulsenergie wurde jedoch im Zentrum eines jeden Intervalls infolge eines Zwei-Photonen-Absorptionsprozesses eines CO_2 -Laser-Photons und des zugehörigen FIR-Photons sowie an den Rändern der Abstimmintervalle beobachtet. Letztere steht im Widerspruch zu theoretischen Rechnungen, die gerade dort einen Anstieg der Kleinsignalverstärkung voraussagen, da dann die Frequenz der CO_2 -Laserstrahlung sehr nahe den Frequenzen der CH_3F -Absorptionslinien liegen. Untersuchungen der Transmission der CH_3F -Ramanzelle für die CO_2 -Laserstrahlung zeigten, daß die Resonanzabsorption dieser Strahlung in den Flanken der CH_3F -Absorptionslinien für die beobachtete Abnahme der FIR-Pulsenergie verantwortlich ist.

In einem ersten Experiment wurde demonstriert, daß das Lasersystem für die lineare Spektroskopie im ferninfraroten Spektralbereich geeignet ist. Dabei wurde als Beispiel das Absorptionsspektrum eines Sauerstoff-dotierten Siliziumkristalls ermittelt. Interstitiell in Silizium eingebauter Sauerstoff führt nämlich zum Auftreten einer resonanten Bandmode, deren Frequenz bei 29.2 cm^{-1} im Frequenzbereich der akustischen Phononen liegt und die durch Absorption von FIR-Strahlung anregbar ist. Es wurde erstmals die genaue spektrale Breite der zugehörigen Absorptionslinie bestimmt. Sie beträgt $(3.9 \pm 0.3) \text{ GHz}$.

Um die Anwendbarkeit des Lasersystems in der nichtlinearen Spektroskopie aufzuzeigen, wurde daraufhin die durch die Anregung der resonanten Bandmode verursachte Absorption bei hohen FIR-Pulsenergien untersucht. Dabei wurde eine starke Zunahme der Transmission eines 5 mm dicken Siliziumkristalls mit einer Sauerstoffkonzentration von $18 \cdot 10^{17} \text{ cm}^{-3}$ mit zunehmender Pulsenergie der FIR-Strahlung beobachtet. Betrug die Energietransmission beispielsweise bei einer FIR-Pulsenergie von $20 \mu\text{J}$ noch 0.2% , erreichte sie bei einer Energie von $200 \mu\text{J}$ bereits einen Wert von 2% . Die Zunahme der Transmission während eines FIR-Laserpulses hoher Energie wurde mit Hilfe zeitaufgelöster Messungen auch direkt beobachtet.

Für die Interpretation der Ergebnisse wurde berücksichtigt, daß sich die resonante Bandmode durch (mittels FIR-Absorption anregbare) Schwingungen der interstitiell im Siliziumkristall eingebauten Sauerstoffatome beschreiben läßt, die mit dem Schwingungssystem des ungestörten Siliziumkristalls durch Emission und Reabsorption resonanter Phononen, d.h. von Phononen mit der gleichen Frequenz wie die Sauerstoffschwingungen, gekoppelt sind. Infolge des durch die interstitielle Einbaulage verursachten anharmonischen Potentials der Sauerstoffschwingungen kann das zugehörige Energieniveausystem bezüglich der FIR-Strahlung von 29.2 cm^{-1} als Zweiniveausystem betrachtet werden. Die Linienbreite der lorentzförmigen FIR-Absorptionslinie zeigt, daß der Energieaustausch zwischen dem System der Sauerstoffschwingungen und dem Schwingungssystem des Kristalls in einer Zeit von 80 psec erfolgt. Da diese Zeit klein gegen die Dauer des FIR-Laserpulses ist, kann davon ausgegangen werden, daß sich die beiden Systeme während des Pulses stets im Gleichgewicht befinden. Eine Modellrechnung wurde daraufhin entwickelt, die die nichtlineare Absorption von FIR-Strahlung durch ein sättigbares System von Zweiniveausystemen beschreibt, das sich stets im Gleichgewicht mit dem nichtsättigbaren System der resonanten Phononen befindet. Der Vergleich der Ergebnisse der Modellrechnung und der experimentell beobachteten Transmission zeigte, daß eine hohe nichtthermische Anregung der resonanten Bandmode zwar für Zeiten existiert, die länger als die Dauer des FIR-Laserpulses sind, daß aber dennoch eine Dämpfung der resonanten Bandmode bereits während des Laserpulses erfolgt. Eine detaillierte Analyse der experimentellen Ergebnisse, insbesondere der zeitlichen Änderung der Transmission während des FIR-Laserpulses, und zusätzlich durchgeführte Experimente zur Ausbreitung der resonanten Phononen ermöglichten den Schluß, daß spektrale Diffusion dieser Phononen aus dem Bereich der Resonanzabsorption den dominanten Zerfallsprozeß der resonanten Bandmode darstellt.

Literaturverzeichnis

- / 1/ V.S. Letokhov und V.P. Chebotayev, *Nonlinear Laser Spectroscopy* (Springer Series in Optical Sciences, Vol.4, Springer Verlag, Berlin 1977).
- / 2/ M.S. Feld und V.S. Letokhov, *Coherent Nonlinear Optics* (Topics in Current Physics, Vol.21, Springer Verlag, Berlin 1980).
- / 3/ W. Böhm, E. Ettlinger und W. Prettl, *Phys.Rev.Lett.* **47**, 1198 (1981).
- / 4/ G.R.Allan, A. Black, C.R. Pidgeon, E. Gornik, W. Seidenbusch und P. Colter, *Phys.Rev. B* **31**, 3560 (1985).
- / 5/ J. Kaminski, J. Spector, W. Prettl und M. Weispfenning, wird veröffentlicht in *Appl.Phys.Lett.*
- / 6/ G.A. Rodriguez, R.M. Hart, A.J. Sievers, F. Keilmann, Z. Schlesinger, S.L. Wright und W.I. Wang, *Appl.Phys.Lett.* **49**, 458 (1986).
- / 7/ T.Y. Chang und T.J. Bridges, *Opt.Comm.* **1**, 423 (1970).
- / 8/ R. Beck, W. Englisch und K. Gürs, *Table of Laser Lines in Gases and Vapors*, 3.Aufl. (Springer Series in Optical Sciences, Vol. 2, Springer Verlag, Berlin 1980).
- / 9/ L.R. Elias, J. Hu und G. Ramian, *Nuclear Instruments and Methods in Physics Research A* **237**, 203 (1985).
- / 10/ A.A. Andronov, I.V. Zverev, V.A. Kozlov, Y.N. Nozdrin, S.A. Pavlov und V.N. Shastin, *JETP Lett.* **40**, 804 (1984).
- / 11/ A.A. Andronov, in *Infrared and Millimeter Waves*, ed. K.J. Button (Academic Press, Inc., Orlando 1986), vol. 16, p. 149.
- / 12/ M. Berg, C.B. Harris, T.W. Kenny und P.L. Richards, *Appl.Phys.Lett.* **47**, 206 (1985).
- / 13/ R. Frey, F. Pradere und J. Ducuing, *Opt.Comm.* **23**, 65 (1977).
- / 14/ A. De Martino, R. Frey und F. Pradere, *Opt.Comm.* **27**, 262 (1978).
- / 15/ D.G. Biron, B.G. Danly, R.J. Temkin und B. Lax, *IEEE J.Quantum Electron.* **QE-17**, 2146 (1981).
- / 16/ P. Mathieu und J.R. Izatt, *Opt.Lett.* **6**, 369 (1981).
- / 17/ B.G. Danly, S.G. Evangelides, R.J. Temkin und B. Lax, *IEEE J.Quantum Electron.* **QE-20**, 834 (1984).
- / 18/ V.N. Bagratashvili, I.N. Knyazev und V.S. Letokhov, *Opt.Comm.* **4**, 154 (1971).
- / 19/ T.Y. Chang und O.R. Wood, *Appl.Phys.Lett.* **23**, 370 (1973).
- / 20/ K. Stenersen und G. Wang, *IEEE J.Quantum Electron.* **QE-19**, 1414 (1983).
- / 21/ F.O'Neill und W.T. Whitney, *Appl.Phys.Lett.* **26**, 454 (1975).

-
- / 22/ V.N. Bagratashvili, I.N. Knyazev, V.S. Letokhov und V.V. Lobko, *Opt.Communic.* **14**, 426 (1975).
 - / 23/ N.G. Basov, E.M. Belenov, V.A. Danilychev, O.M. Kerimov, I.B. Kovsh, A.S. Podsonnyi und A.F. Suchkov, *Sov.Phys.-JETP* **37**, 58 (1973).
 - / 24/ N.W. Harris, F. O'Neill und W.T. Whitney, *Opt.Communic.* **16**, 57 (1976).
 - / 25/ A.J. Alcock, K. Leopold und M.C. Richardson, *Appl.Phys.Lett.* **23**, 562 (1973).
 - / 26/ D. Rollin und A.J. Alcock, *Opt.Communic.* **23**, 11 (1977).
 - / 27/ T.W. Carman und P.E. Dyer, *Opt.Communic.* **29**, 218 (1979).
 - / 28/ P.E. Dyer und B.L. Tait, *Appl.Phys.Lett.* **37**, 356 (1980).
 - / 29/ B.K. Deka, P.E. Dyer und R.J. Winfield, *Opt.Communic.* **39**, 255 (1981).
 - / 30/ B.K. Deka, P.E. Dyer und R.J. Winfield, *Opt.Lett.* **5**, 194 (1980).
 - / 31/ B.K. Deka, P.E. Dyer und R.J. Winfield, *Opt.Communic.* **33**, 206 (1980).
 - / 32/ A.J. Alcock und P.B. Corkum, *Phil.Trans.Roy.Soc.Lond. A* **298**, 365 (1980).
 - / 33/ P.B. Corkum, *IEEE J.Quantum Electron.* **QE-21**, 216 (1985).
 - / 34/ Wan Chong-Yi, W. Fuss, C. Schwab, Zhou Jin-Wen und K.L. Kompa, Report MPQ 86, Max-Planck-Institut für Quantenoptik, Garching 1984.
 - / 35/ T. Jaeger und G.Wang, in *Tunable Lasers*, ed. L.F. Mollenauer und J.C. White (Topics in Applied Physics, Vol. 59, Springer Verlag, Heidelberg 1987), p. 303.
 - / 36/ G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules* (Van Nostrand Reinhold Comp., New York 1945).
 - / 37/ W.J. Wittenman, *The CO₂ Laser* (Springer Series in Optical Sciences, Vol. 53, Springer Verlag, Berlin 1987).
 - / 38/ O.R. Wood, *Proceedings of the IEEE* **62**, 355 (1974).
 - / 39/ J.T. Verdeyen, *Laser Electronics* (Prentice Hall, Inc., Englewood Cliffs, New Jersey 1981).
 - / 40/ N.C. Chang und M.T. Tavis, *IEEE J.Quantum Electron.* **QE-10**, 372 (1974).
 - / 41/ J. Reid, J.Shewchun und B.K. Garside, *Appl.Phys.* **17**, 349 (1978).
 - / 42/ R.L. Abrams, *Appl.Phys. Lett.* **25**, 609 (1974).
 - / 43/ J.L. Miller, *J.Appl.Phys.* **49**, 3076 (1978).
 - / 44/ W.E. Schmid, Report IPP IV/84, Max-Planck-Institut für Plasmaphysik, Garching 1975.
 - / 45/ I. Burak, Y. Noter und A. Szöke, *IEEE J.Quantum Electron.* **QE-9**, 541 (1973).
 - / 46/ E.A. Ballik, B.K. Garside, J. Reid und T. Tricker, *J.Appl.Phys.* **46**, 1322 (1975).
 - / 47/ R.S. Taylor, A.J. Alcock, W.J. Sarjeant und K.E. Leopold, *IEEE J.Quantum Electron.* **QE-15**, 1131 (1979).
 - / 48/ S. Lovold und G. Wang, *IEEE J.Quantum Electron.* **QE-20**, 182 (1984).
 - / 49/ E. Nasser, *Fundamentals of Gaseous Ionization and Plasma Electronics* (Wiley-Interscience, New York 1971).

-
- / 50/ J.J. Lowke, A.V. Phelps und B.W. Irwin, J.Appl.Phys. **44**, 4664 (1973).
 - / 51/ L.J. Denes und J.J. Lowke, Appl.Phys.Lett. **23**, 130 (1973).
 - / 52/ W.L. Nighan und J.H. Bennett, Appl.Phys.Lett. **14**, 240 (1969).
 - / 53/ J.I. Levatter und Lin Shao-chi, J.Appl.Phys. **51**, 210 (1980).
 - / 54/ A.J. Palmer, Appl.Phys.Lett. **25**, 138 (1974).
 - / 55/ V.N. Karnyushin, A.N. Malov and R.I. Soloukhin, Sov.J.Quantum.Electron. **8**, 319 (1978).
 - / 56/ D.C. McKen, H.J. Seguin und J. Tulip, IEEE J.Quantum Electron. **QE-12**, 470 (1976).
 - / 57/ P. Mathieu, Dissertation, Universität Laval, Kanada (1982).
 - / 58/ Wan Chong-Yi, C. Schwab, W. Fuss und K.L. Kompa, Opt.Communic. **46**, 311 (1983).
 - / 59/ H. Shields, A.L.S. Smith und B. Norris, J.Phys. D : Appl.Phys. **9**, 1587 (1976).
 - / 60/ R.B. Gibson, A. Javan und K. Boyer, Appl.Phys.Lett. **32**, 726 (1978).
 - / 61/ Wan Chong-Yi, U. Werling und K.F. Renk, J.Appl.Phys. **57**, 990 (1985).
 - / 62/ U. Werling, Wan Chong-Yi und K.F. Renk, Intern.J. Infrared and Millimeter Waves **6**, 449 (1985).
 - / 63/ B.G. Danly, S.G. Evangelides, R.J. Temkin und B. Lax, in *Infrared and Millimeter Waves*, vol. 12, ed. K.J. Button (Academic Press, Inc., New York 1984), p. 195.
 - / 64/ T.Y. Chang, Rev.Sci.Instrum. **44**, 405 (1973).
 - / 65/ U. Tietze und Ch. Schenk, *Halbleiterschaltungstechnik* (Springer Verlag, Berlin 1983).
 - / 66/ G. Busse und R. Thurmaier, Appl.Phys.Lett. **31**, 194 (1977).
 - / 67/ W.L. Smith und I.M. Mills, J.Mol.Spectrosc. **11**, 11 (1963).
 - / 68/ W. Eisfeld, U. Werling und W. Prettl, Appl.Phys.Lett. **42**, 276 (1983).
 - / 69/ K. Moser, W. Eisfeld, U. Werling, S. Wahl und W. Prettl, Appl.Phys.Lett. **45**, 711 (1984).
 - / 70/ J. Gilbert und J.L. Lachambre, Appl.Phys.Lett. **18**, 187 (1971).
 - / 71/ H. Statz und C.L. Tang, J.Appl.Phys. **36**, 3924 (1965).
 - / 72/ V.I. Malyshev, A.S. Markin, A.V. Masalov und A.A. Sychev, Soviet Physics JETP **30**, 453 (1970).
 - / 73/ P. Bernard, P. Mathieu und J.R. Izatt, Opt.Communic. **37**, 285 (1981).
 - / 74/ B.S. Patel, Appl.Optics **16**, 1232 (1977).
 - / 75/ J.R. Izatt und B.K. Deka, Int.J.Infrared and Millimeter Waves **5**, 1473 (1984).
 - / 76/ U. Werling, K.F. Renk und Wan Chong-Yi, Intern.J.Infrared and Millimeter Waves **7**, 881 (1986).
 - / 77/ U. Werling, Wan Chong-Yi und K.F. Renk, IEEE J. Quantum Electron. **QE-22**, 1024 (1986).
 - / 78/ P. Shoja-Chaghervand und R.H. Schwendeman, J.Mol.Spectrosc. **98**, 27 (1983).

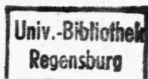
-
- / 79/ N.J. Trappeniers und E.W.A. Elenbaas-Bunschoten, *Chem.Phys.Lett.* **64**, 205 (1979).
 - / 80/ E. Weitz, G. Flynn und A.M. Ronn, *J.Chem.Phys.* **56**, 6060 (1972).
 - / 81/ R. Loudon, *The Quantum Theory of Light* (Clarendon Press, Oxford 1983).
 - / 82/ D. Marcuse, *Principles of Quantum Electronics* (Academic Press, New York 1980).
 - / 83/ D.R. Hall, E.K. Gorton und R.M. Jenkins, *J.Appl.Phys.* **48**, 1212 (1977).
 - / 84/ C.T. Gross, J. Kiess, A. Mayer und F. Keilmann, *IEEE J. Quantum Electron.* **QE-23**, 377 (1987).
 - / 85/ K.F. Renk und L. Genzel, *Appl.Optics* **1**, 643 (1962).
 - / 86/ R. Ulrich, K.F. Renk und L. Genzel, *IEEE Trans. on Microwave Theory and Techniques* **11**, 363 (1963).
 - / 87/ L.O. Hocker, A. Javan, R. Ramachandra Rao, L. Frenkel und T. Sullivan, *Appl.Phys. Lett.* **10**, 147 (1967).
 - / 88/ J.R. Izatt, B.K. Deka und Wen-Sen Zhu, *IEEE J.Quantum Electron.* **QE-23**, 117 (1987).
 - / 89/ E. Kalikstein, Diplomarbeit, Universität Regensburg (1986).
 - / 90/ *Semiconductor Silicon 1981*, ed. by H.R. Huff, R.J. Kriegler und Y. Takeishi (Electrochemical Soc., Pennington, 1981).
 - / 91/ *Defects in Semiconductors*, ed. by J. Narayan und T.Y. Tan (North-Holland, New York 1981).
 - / 92/ U. Gösele und T.Y. Tan, *Appl.Phys. A* **28**, 79 (1982).
 - / 93/ L. Jastrzebski, *IEEE Trans. on Electron. Devices* **ED-29**, 475 (1982).
 - / 94/ G.S. Oehrlein und J.W. Corbett, in *Defects in Semiconductors II*, ed. by S. Mahajan und J.W. Corbett (North-Holland, New York 1983), p. 107.
 - / 95/ W.L. Bond und W. Kaiser, *J.Phys.Chem.Solids* **16**, 44 (1960).
 - / 96/ R.C. Newman, *Infra-red Studies of Crystal Defects* (Taylor & Francis Ltd. London 1973).
 - / 97/ H.J. Hrostowski und R.H. Kaiser, *J.Phys.Chem.Solids* **9**, 214 (1959).
 - / 98/ R.M. Chrenko, R.S. McDonald und E.M. Pell, *Phys.Rev.* **138**, A1775 (1965).
 - / 99/ Landolt-Börnstein, *Zahlenwerte und Funktionen aus Naturwissenschaft und Technik*, Gruppe III, Bd. 17a (Springer Verlag, Heidelberg 1982).
 - / 100/ D.R. Bosomworth, W. Hayes, A.R.L. Spray und G.D. Watkins, *Proc.Roy.Soc.Lond.* **A 317**, 133 (1970).
 - / 101/ W. Eisenmenger, in *Nonequilibrium Superconductivity, Phonons and Kapitza Boundaries*, ed. K.E. Gray, NATO ASI Series, Vol. B 65 (Plenum Press, New York 1981), p. 73.
 - / 102/ H.J. Hrostowski und B.J. Alder, *J.Chem.Phys.* **33**, 980 (1960).
 - / 103/ W.R. Thorson und I. Nakagawa, *J.Chem.Phys.* **33**, 994 (1960).

-
- /104/ J.R. Aronson, R.C. Lord und D.W. Robinson, J.Chem.Phys. **33**, 1004 (1960).
- /105/ M. Stavola, Appl.Phys.Lett. **44**, 514 (1984).
- /106/ T. Iizuka, S. Takasu, M. Tajima, T. Arai, T. Nozaki, N. Inoue und M. Watanabe, J.Electrochem.Soc. **132**, 1707 (1985).
- /107/ H.Bilz und W.Kress, in *Phonon Dispersion Relations in Insulators*, (Springer Series in Solid-State Sciences, Vol.10, Springer Verlag, Heidelberg 1979), p. 97.
- /108/ L. Genzel, in *Optical Properties of Solids*, ed. by S. Nudelman und S.S. Mitra (Plenum Press, New York 1969), p. 453.
- /109/ A.A. Maradudin, in *Solid State Physics*, ed. F. Seitz und D. Turnbull, (Academic Press, New York 1966), Vol. 18, p. 273; Vol. 19, p. 1, Suppl. 3, p. 353.
- /110/ A.S. Barker, Jr. und A.J. Sievers, Rev. Mod. Phys. **47**, Suppl. No. 2, S1 (1975).
- /111/ H. Bilz, D. Strauch und R.K. Wehner, in *Handbuch der Physik*, Vol. 25/2d, ed. L. Genzel (Springer Verlag, Berlin 1984), p. 262.
- /112/ M.v. Klein, in *Physics of Colour Centers*, ed. W.B. Fowler (Academic Press, New York 1968), p. 430.
- /113/ S. Takeno, Prog.Theor.Phys. **29**, 191 (1963).
- /114/ D.K. Brice, Phys.Rev. **140**, A1211 (1965).
- /115/ M.Wagner, Phys.Rev. **133**, A750 (1964).
- /116/ P.H. Dederichs und R. Zeller, Phys.Rev. B **14**, 2314 (1976); R. Zeller und P.H. Dederichs, Zeitschrift f. Physik B **25**, 139 (1976).
- /117/ Y.B. Levinson, in *Nonequilibrium Phonons in Nonmetallic Crystals*, ed. W. Eisenmenger und A.A. Kaplyanskii (North-Holland, Amsterdam 1986), p. 91.
- /118/ K.F. Renk, in *Nonequilibrium Phonons in Nonmetallic Crystals*, ed. W. Eisenmenger und A.A. Kaplyanskii (North-Holland, Amsterdam 1986), p. 317.
- /119/ R.H. Pantell und H.E. Puthoff, *Fundamentals of Quantum Electronics* (John Wiley & Sons, New York 1969).
- /120/ E.V. Loewenstein, D.R. Smith und R.L. Morgan, Appl. Optics **12**, 398 (1973).
- /121/ L. Allen und J.H. Eberly, *Optical Resonance and Two-Level Atoms* (John Wiley & Sons, New York 1975).
- /122/ J. Crank, *The Mathematics of Diffusion* (Clarendon Press, Oxford 1979).
- /123/ U. Werling, Diplomarbeit, Universität Regensburg (1982).
- /124/ R.W. Alexander, Jr., A.E. Hughes und A.J. Sievers, Phys.Rev. B **1**, 1563 (1970).
- /125/ A. Berke, A.P. Mayer und R.K. Wehner, Solid State Commun. **54**, 395 (1985).
- /126/ S. Tamura, Phys.Rev. B **27**, 858 (1973).
- /127/ M.K. Maul und M.W.P. Strandberg, J.Appl.Phys. **40**, 2822 (1969).
- /128/ W. Eisfeld, Dissertation, Universität Regensburg (1979).
- /129/ U. Happek, T. Holstein und K.F. Renk, Phys.Rev.Lett. **54**, 2091 (1985).
- /130/ U. Happek, T. Holstein und K.F. Renk, Phys.Rev. B **34**, 8898 (1986).

-
- /131/ N. Retzer, Dissertation, Universität Regensburg (1987).
/132/ U. Werling und K.F. Renk, Japanese J.Appl.Phys. **26**, 671 (1987).
/133/ E.I. Gershenzon, G.N. Goltsman und A.P. Melnikov, JETP Letters **14**, 185 (1971).
/134/ W. Burger, Dissertation, Universität Stuttgart (1986).
/135/ T.M. Lifshits und F.Ya. Nad, Sov.Phys. - Dokl. **10**, 532 (1965).
/136/ H. Lengfellner, U. Werling, J. Hummel, H. Netter, N. Retzer und K.F. Renk, Solid State Commun. **38**, 1215 (1981).
/137/ A.A. Volkov, Yu.G. Goncharov, G.V. Kozlov, S.P. Lebedev und A.M. Prokhorov, Infrared Phys. **25**, 369 (1985).
/138/ D. Grandjean, J. Claudel, F. Bréhat, A. Hadni, P. Strimer und R. Thomas, J. Phys. (France) **31**, 471 (1970).
/139/ A. Hadni und X. Gerbeaux, Opt.Comm. **50**, 231 (1984).
/140/ L.C. Lee und W.L. Faust, Phys.Rev.Lett. **26**, 648 (1971).

An dieser Stelle möchte ich mich bei all denen bedanken, die zum Gelingen dieser Arbeit beigetragen haben, besonders bei

- Herrn Prof. Dr. K.F. Renk für die interessante Aufgabenstellung, die vielen kritischen Diskussionen und die großzügige materielle Unterstützung;
- Herrn Prof. Wan Chong-Yi, Institute of Electronics, Academia Sinica in Peking, China, der die Konstruktionsvorschläge für den Hochdruck-CO₂-Laser lieferte und der mich mit den Grundlagen der Hochspannungs-Gasentladungstechnik vertraut gemacht hat;
- Herrn Prof. Dr. W. Prettl für die ständige Diskussionsbereitschaft;
- Herrn E. Kalikstein, der im Rahmen seiner Diplomarbeit den Meßplatz für die Transmissionsspektroskopie optimierte und der eine große Hilfe bei der Beseitigung der technischen Probleme des Hochdruck-CO₂-Lasers war;
- Herrn Dr. H. Lengfellner für seine Tips zur Herstellung supraleitender Bolometer;
- meinen Kollegen Dr. H. Schwartz, Dr. E. Ettlinger, U. Happek, G. Jungwirt, P.E. Obermayer, Dr. N. Retzer und M. Weispfenning, die bei allen anfallenden Schwierigkeiten diskussions- und hilfsbereit waren;
- allen übrigen Mitgliedern des Lehrstuhls für die freundschaftliche Zusammenarbeit;
- Frau M. Reif für die graphische Gestaltung dieser Arbeit und ihre sonstige stetige Hilfsbereitschaft;
- den Herren A. Humbs und K. Lachner für ihren Einsatz und ihre technische Unterstützung während der Aufbauphase des Lasersystems;
- Herrn F. Waider, stellvertretend für die feinmechanische Werkstatt der Fakultät Physik, in der alle Teile des mechanischen Aufbaus des Lasersystems gefertigt wurden und die bei kurzfristig aufgetretenen Problemen stets mit Rat und Tat zur Seite stand;
- den Herren K. Walling und K.H. Weigert für die stetige Bereitstellung von flüssigem Helium;
- Herrn G.Göbl und seinen Mitarbeitern in der Elektronikwerkstatt für ihre Hilfe bei elektronischen Problemen;
- den Herren Dr. J. Reffle und Dr. D. Huber der Firma Wacker-Chemitronic, Burghausen, bzw. Herrn Dr. K. Laßmann, Universität Stuttgart, die freundlicherweise Sauerstoff-dotierte Siliziumkristalle zur Verfügung stellten.



ubr09070



00/QC 130 F828

Frank, Maximilian : Kooperation und Koordination : eine spieltheoretische Interpretation klassischer Marktversagungsgründe. - Weiden ; Regensburg : eurotrans-Verl., 1996

Medien-Nr.: 069020952526
BV-Nr.: BV010926101
gedruckt: 02.07.2011 15:05:08
Status: angelegt

Projekt: Regensburger Dissertationen

Maximilian Frank

Kooperation und Koordination

**Eine spieltheoretische Interpretation klassischer
Marktversagungsgründe**

**00
QC
130
F828**

eurotrans-Verlag

UBR069020952526



Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft, eingereicht
an der Wirtschaftswissenschaftlichen Fakultät der Universität Regensburg

Berichterstatter: Prof. Dr. Wolfgang Buchholz
Prof. Dr. Winfried Vogt

Tag der Disputation: 22.12.1994

Maximilian Frank

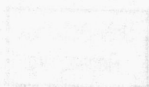
Kooperation und Koordination

Eine spieltheoretische Interpretation
klassischer Marktversagungsgründe

Maximilian Frank

Kooperation und Koordination

Eine spieltheoretische Interpretation klassischer
Marktversagungsgründe



eurotrans-Verlag

0196/0335

00 / 6C 130 - F 828

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Frank, Maximilian:

Kooperation und Koordination : eine spieltheoretische Interpretation klassischer Marktversagungsgründe / Maximilian Frank. - Weiden ; Regensburg : eurotrans-Verl., 1996

Zugl.: Regensburg, Univ., Diss., 1994

ISBN 3-929318-45-8

Univ.-Bibliothek
Regensburg

109261011

ISBN 3-929318-45-8

Alle Rechte vorbehalten

© 1996, eurotrans-Verlag, Weiden und Regensburg

Druck: Gräbner, Bamberg

Printed in Germany

INHALT

Näherung	3
----------	---

Zur Theorie des Marktversagens

1	Soziale Dilemmata und Externalitäten	8
1.1	Eine Typologie sozialer Dilemmata	11
1.2	Externalitäten	20
2	Auswege ?	30
3	Verhandlungen und Bindung - eine Akzentuierung	34

Versuch über die Idee der Reputation

(Zur Begründung von Kooperation in einem endlich wiederholten Spiel)

1	Reputation und wiederholte Spiele	45
2	Das Grundmodell	47
2.1	Ein Poolgleichgewicht	50
2.2	Ein Separationsgleichgewicht	53
3	Variation 1: Unsicherheit über die pay-offs der Egoistin	59
4	Variation 2: Zwei Egoistentypen	66
5	Variation 3: Ein n-Perioden Modell	69

Versuch über die Idee der Bindung

(Zum Risiko einer commitment-Strategie in einem Problem der Koordination)

1	Bindung und Risiko	75
2	Ein Beispiel aus der Umweltökonomie	77
3	Fairness	85

Ausklang	93
----------	----

Anhang	95
--------	----

Literatur	99
-----------	----

Der Gegenstand, von dem ich hier handle¹, ist die Frage der Kooperation und der Koordination. Diese Frage erwächst aus einem Interesse an Konflikten, Konflikten zwischen Menschen, zwischen Institutionen oder auch zwischen Nationen. Dieses Thema läßt sich naturgemäß niemals in allen Ausprägungen und Betrachtungsweisen erfassen. Aus diesem Grund war es zwingend, eine Auswahl zu treffen, sowohl hinsichtlich der inhaltlichen Gedankenführung als auch hinsichtlich des Stils der Präsentation.

Näherung und Ausklang umschließen diese Abhandlung. In ihr finden sich eine Fühlungnahme mit dem Thema ("Zur Theorie des Marktversagens") sowie zwei Modellgruppen, die sich zum einen der Frage der Reputation ("Versuch I"), zum anderen der Frage der Bindung ("Versuch II") widmen. Der Aufbau jedes einzelnen Kapitels unterliegt einer weiteren Strukturierung. Betrachten wir die Fühlungnahme:

I

Alles beginnt mit dem Konflikt. Dadurch sind wir sofort bei der Frage des *Stils*. Um diesen Zusammenhang zu verdeutlichen, sei auf eine Äußerung von Igor Strawinsky aus seinen Harvard-Vorlesungen zur musikalischen Poetik verwiesen:

"Was mich betrifft, so überläuft mich eine Art von Schrecken, wenn ich im Augenblick, wo ich mich an die Arbeit begeben, die unendliche Zahl der mir sich bietenden Möglichkeiten erkenne und fühle, daß mir alles erlaubt ist. Wenn mir alles erlaubt ist, das Beste und das Schlimmste, wenn mir nichts Widerstand bietet, dann ist jede Anstrengung undenkbar, ich kann auf nichts bauen, und jede Bemühung ist demzufolge vergebens."

Igor Strawinsky (1983), S.212

¹ Die Entlehnung des eröffnenden Satzes bei Martin Buber (1994) sei nicht verschwiegen.

Die Typologie sozialer Dilemmata (Abschnitt 1.1) charakterisiert mit Hilfe einer einfachen spieltheoretischen Darstellung unterschiedliche Arten eines Konflikts, auf der einen Seite Konflikte, denen ein Kooperationsproblem zugrundeliegt, auf der anderen Seite Konflikte, die sich auf die Schwierigkeit der Koordination zurückführen lassen. Da ich zu dieser Charakterisierung ein einheitliches Modell verwende, offenbare ich die beiden Bestandteile meiner Stilistik: Ordnung der Mannigfaltigkeit der Erscheinungen durch Analogie und Kontrast². Die Sprache, die ich zu diesem Zweck wähle, ist über weite Strecken spieltheoretisch motiviert, beinahe ausschließlich gilt dies in den beiden Versuchen.

Nach der Vorstellung des Themas begründe ich dessen Relevanz in Anknüpfung an die (neo)klassische Theorie des Marktversagens (Abschnitt 1.2). Externalitäten betrachte ich in all ihren Variationen als Grundlage für ein Versagen des marktwirtschaftlichen Systems, Externalitäten münden in die Frage der Kooperation oder in die Frage der Koordination.

Es folgt der zweite Abschnitt. Bereits vorliegende Ansätze können durch die Variation der Komponenten eines neoklassischen Modells unterschieden werden: Variation der Zielfunktion, Variation der Spielstruktur oder Variation der Restriktion. Bei allen Unterschieden kommen dabei die gemeinsamen, tragenden Motive zum Vorschein: Erwartungen, Glaubwürdigkeit und Bindung.

Im dritten Abschnitt greife ich den Gedanken auf, der meiner Modellierung des zweiten "Versuchs" zugrundeliegt: Bindung durch selbstgewählte Beschränkung der Handlungsmöglichkeiten.

Im Mittelpunkt dieser "Akzentuierung" steht die Frage der Akzeptanz einer Bindung. In ihr unterscheide ich zuerst zwischen exogenen und endogenen Restriktionen, sowie innerhalb letzterer zwischen inter- und intrapersonellen Überlegungen. Daran anknüpfend wende ich mich Konflikten zu, die auf der Teilung einer gegebenen Ressource beruhen. Für die Einigung auf eine bestimmte Aufteilung ist die Akzeptanz einer Teilungsregel notwendig. Die beteiligten Individuen müssen sich daher binden. In der volkswirtschaftlichen Literatur findet sich in der Theorie der Verhandlungen das methodische Instrumentarium zur Untersuchung dieser Fragestellungen. Bindung erscheint in dieser Theorie in zwei Gestalten: die erste Vorstellung beinhaltet die Akzeptanz einer axiomatisch begründeten Teilungsfunktion (axiomatische Verhandlungstheorie), die zweite Vorstellung die Akzeptanz eines institutionellen Rahmens, in dem der Verhandlungsprozeß stattfindet (strategische

² Wiederum: Igor Strawinsky (1983). Kontrapunkt: Paul Feyerabend (1983).

Verhandlungstheorie). Neben einer Präsentation der jeweiligen Standardmodelle aus diesen beiden Richtungen gebe ich in einzelnen Anmerkungen verschiedene Konsequenzen und Zusammenhänge zu bedenken. Mit der Rückkehr zum Gedanken der Bindung beende ich diesen Abschnitt.

II

Die Bildung von Reputation als Ursache für die Entstehung von Kooperation war die Leitlinie bei der Entwicklung des ersten "Versuchs". Würde es gelingen, sogar in einer Welt mit endlichem Zeithorizont in einer dem Gefangenendilemma sehr ähnlichen Situation einen Anreiz für kooperatives Verhalten zu begründen?

Die gesamte Abhandlung, nicht nur das an dieser Stelle präsentierte Modell, ist von der Relevanz asymmetrischer oder unvollkommener Information durchdrungen. In ihr liegt oftmals die Ursache für die Existenz eines sozialen Dilemmas, hier wird sie zur Grundlage eines Auswegs. Wohlan, wenden wir uns spieltheoretischer Analyse und damit spieltheoretischer Sprache zu³.

Das Hauptargument findet sich im Signalcharakter der Aktion der informierten Spielerin I: die Wahrscheinlichkeitsverteilung der Spielerin II über die möglichen Typen von Spielerin I zu einem Zeitpunkt i wird durch die früheren Aktionen der Spielerin I beeinflusst. Der Typus von Spielerin I wird durch ihre Präferenzordnung charakterisiert, als Altruistin oder Egoistin. Es wird das Konzept eines sequentiellen Gleichgewichts angewendet. Das Grundmodell entwickelt das Ergebnis im einfachst möglichen Rahmen (Abschnitt 2). Dazu war es an einer Stelle notwendig, kontrafaktische Vermutungen bei Spielerin II zu unterstellen. Diese harte und wohl kaum zu verteidigende Annahme wird in den folgenden Variationen aufgegeben.

Abschnitt 3 modelliert Unsicherheit bei Spielerin II über die pay-offs von Spielerin I, Abschnitt 4 interpretiert die entsprechenden Vermutungen nicht mehr als subjektive Wahrscheinlichkeiten, sondern als existierende Anteile unterschiedlicher Typen von Egoistinnen.

³ Bemerkungen über die Spieltheorie finden sich bei fast allen Spieltheoretikern. Als (subjektive) Auswahl sei hier nur auf einige Veröffentlichungen verwiesen: Aumann (1985), Binmore (1990), Kreps (1990c) und Rubinstein (1991, 1992).

In den Abschnitten 2 - 4 ist die Spieldauer auf zwei Perioden begrenzt. In Abschnitt 5 wird dagegen die Betrachtung auf n Perioden ausgeweitet.

Im Ergebnis kooperiert eine Egoistin, obwohl Nichtkooperation in jeder einzelnen Runde dominante Strategie ist, in $n - 1$ Perioden. Der Grund liegt im wesentlichen darin, daß sie dadurch bei ihrer Gegenspielerin die Unsicherheit darüber, daß sie auch eine Altruistin sein könnte, am Leben erhält. Die Frage, ob es sinnvoll ist, eine Egoistin mit diesem Verhalten weiterhin als Egoistin zu bezeichnen, möchte ich hier nicht beantworten, sondern nur stellen.

III

Bindung, angesprochen im dritten Abschnitt des ersten Teils ("Zur Theorie des Marktversagens"), ist das Motiv des zweiten "Versuchs". Es wird angewendet auf eine Ausprägung des Koordinationsproblems, es wird entwickelt in der freiwilligen Selbstbindung durch Verzicht auf eine Handlungsoption, es wird verbunden mit unvollkommener (Abschnitt 2) bzw. asymmetrischer (Abschnitt 3) Information.

In den beiden Hauptabschnitten 2 und 3 verwende ich zwei unterschiedliche Arten inhaltlicher Motivation. Ich beginne mit der Vorstellung von Abschnitt 2:

Internationale Verhandlungen zur Reduktion von CO_2 werden unter Berücksichtigung globaler Klimamodelle geführt, deren Prognosen bzgl. der Konsequenzen eines weiteren CO_2 -Anstiegs einer erheblichen Varianz unterliegen. Die Strategie eines Landes, sich beispielsweise durch ihre Energiepolitik an "Nichtkooperation" in einem globalen Spiel der Begrenzung von CO_2 zu binden, birgt gravierende Risiken. Dies gilt vor allem, wenn man zusätzlich die Irreversibilität bereits emittierter Treibhausgase bedenkt.

Das Modell wurde sehr einfach gewählt. Es besteht aus zwei Stufen. In der ersten Stufe besteht für einen Spieler die Möglichkeit, sich an "Nichtkooperation" zu binden. Die Gestalt des Spiels in der zweiten Stufe ist von dieser Entscheidung abhängig. Im ersten Schritt wird bei Sicherheit argumentiert, im zweiten Schritt führe ich Unsicherheit über die Nutzenwerte ein.

Der dritte Abschnitt beinhaltet den Gedanken der Fairness. Er argumentiert in der Tradition der Vorwärtsinduktion: die Bindungsentscheidung eines Spielers beeinflusst die Präferenzen des anderen Spielers. Verzichtet der erste Spieler auf eine mögliche Bindung, wird diese Aktion vom zweiten Spieler als fair erachtet, im anderen Fall

umgekehrt. Wieder gehe ich zuerst von einem Modell bei Sicherheit aus, anschließend diskutiere ich den Fall der Unsicherheit beim ersten Spieler über die Höhe der Fairnessempfindungen des zweiten Spielers.

Bevor wir mit dem ersten Kapitel beginnen, sei mir noch eine Anmerkung gestattet: diese Arbeit entstand in enger Zusammenarbeit mit den Mitgliedern des Institutes für Volkswirtschaftslehre der Universität Regensburg.

Meine beiden Doktorväter, Wolfgang Buchholz und Winfried Vogt, begleiteten auf vielfältige Weise. Sie waren Anstoßgeber, sie haben stets kritisch hinterfragt, sie waren offen und aufgeschlossen auch gegenüber eher unkonventionelleren Fragestellungen oder Darstellungen. Sie haben mich immer wieder unterstützt und neue Wege aufgezeigt. Vor allem aber: sie sind Menschen von Intellekt *und* Herz.

Sehr viele der auf den folgenden Seiten präsentierten Gedanken entsprangen aus Diskussionen mit den Professoren und mit meinen ehemaligen Kollegen am Institut. Ihre Anregungen, ihre Zweifel, ihr Interesse und ihre Kritik haben diese Abhandlung Gestalt annehmen lassen. Mein Dank gilt daher allen Mitgliedern des Institutes - im besonderen Wilhelm Althammer, Georg Götz, Joachim Grosser, Christian Haslbeck und Claudia Löhnig.

Erwähnen möchte ich an dieser Stelle ebenso Thomas Schmid-Schönbein. Während der Erstellung meiner Diplomarbeit im Herbst 1988 führten wir lange Gespräche über die Erwartungsnutzenhypothese, über Rationalität, über die neoklassische Methodik und über die Struktur wissenschaftlicher Gedankenführung. Diese Dialoge spiegeln sich auch in diesen Seiten wider.

1. Soziale Dilemmata und Externalitäten

Die Geschichte ökonomischer Theorie erstreckt sich auf einen vergleichsweise kurzen Zeitraum, verglichen mit der Geschichte der Staatstheorien. Die zentrale Frage, das Verhältnis von Staat und Individuum, die Rechtfertigung staatlicher Eingriffe in die Freiheitsrechte des Einzelnen, ist beiden gemein. Gleichwohl finden sich in der ökonomischen Theorie spezifische Akzente: die Entscheidung *zwischen* Wirtschaftssystemen als auch die Frage nach der Wirtschaftspolitik *innerhalb* eines gegebenen Wirtschaftssystems. Die politischen Ereignisse Ende der 80er und zu Beginn der 90er Jahre dieses Jahrhunderts haben - zumindest in einem weiten Bereich - das Interesse an der ersten Fragestellung auf den Aspekt der Transformation reduziert.

In einem idealtypischen marktwirtschaftlichen System besteht die Minimalfunktion des Staates aus der Sicherung der Eigentumsrechte. Daneben wird dieses System durch die Existenz nicht-kontrollierter Preise und freier Märkte charakterisiert. Um darüber hinaus staatliche Eingriffe begründen zu können, bedarf es des Nachweises von Marktfehlern. Dazu ist es notwendig, sich die Voraussetzungen einer idealtypischen Marktwirtschaft zu vergegenwärtigen. Das grundlegende Modell, dessen Basis durch die Arbeit von Gerard Debreu (1959) gelegt wurde, ist die allgemeine Gleichgewichtstheorie⁴. Die bedeutendsten Annahmen dieses Modells sind: eine kostenlose und vollständige Spezifikation von Eigentum, kostenlose Einrichtung und Betrieb von Märkten (keine Marktzutrittsschranken und keine Transaktionskosten), kostenlose und symmetrisch verteilte Information sowie fehlende Marktmacht der Akteure. Ist irgendeine dieser Annahmen verletzt, spricht man von Marktversagen und begründet damit einen staatlichen Eingriff.

⁴ Als Input fließen in dieses Modell die Präferenzen der Akteure, ihre Anfangsausstattungen an Produktionsfaktoren sowie die vorherrschende Technologie, als Ergebnis resultiert ein eindeutiges und stabiles Gleichgewicht. Als klassische Referenz neben Debreu gilt Arrow/Hahn (1971).

Diese Abhandlung orientiert sich an der These, daß die meisten Marktfehler durch ein Kooperations- oder ein Koordinationsproblem charakterisiert werden können. Die beiden Probleme treten im wesentlichen dann auf, wenn die Bürger die Konsequenzen ihrer Handlungen auf ihre Mitmenschen nur unvollkommen berücksichtigen und von der Vermutung ausgehen, daß dies für die jeweils anderen ebenso gilt. Die eigene Verhaltensänderung *und* die entsprechende Erwartungsbildung über das Verhalten der anderen sind die entscheidenden Komponenten eines Auswegs aus diesen Problemen. Eine Lösung kann entweder in traditioneller Weise durch einen Staatseingriff versucht werden oder aber durch privates Handeln⁵.

Damit wird neben der politischen auch die soziale Dimension dieser Fragestellung erkennbar. Die Stabilität einer Gesellschaft wird wesentlich von ihrer Fähigkeit bestimmt, soziale Dilemmata in Form von Kooperations- oder Koordinationsproblemen zu lösen. Die unterschiedlichen Wege hierzu prägen das Gesicht der jeweiligen Gesellschaft. Die Verbindung dieser Überlegungen zur ökonomischen Theorie verdeutlicht Jon Elster in der Einleitung seines Buches "The cement of society":

"I shall discuss two concepts of social order: that of stable, regular, predictable patterns of behaviour and that of cooperative behaviour. ... Instead of referring to predictability and cooperation, economists talk about equilibrium and Pareto optimality."

Jon Elster (1989), S.1

Mit diesem Zitat möchte ich die einführenden Bemerkungen beschließen. In Abschnitt 1.1 beginne ich mit einer Typologie möglicher Ausprägungen eines Kooperations- oder Koordinationsproblems anhand eines einfachen 2-Personen-Spiels. Anschließend (Abschnitt 1.2) wird diskutiert, inwieweit sich die traditionellen Marktversagensgründe (öffentliche Güter, asymmetrische Informationen, zunehmende Skalenerträge) als verschiedene Formen eines externen Effekts begreifen lassen. In der weiteren Argumentation verweise ich auf den Zusammenhang zwischen Externalitäten und den verschiedenen Formen sozialer Dilemmata. Im zweiten Abschnitt werden die zentralen

⁵ Es mag in diesem Zusammenhang angemessener sein, zwischen exogener und endogener Lösung zu unterscheiden, wobei die Grenzen fließend sein können. Der Staat oder auch andere Institutionen (man denke beispielsweise an die Kirche) greifen in die Entscheidungen des Einzelnen u.a. durch die Etablierung eines Rechtssystems ein. Wird ein exogen gesetztes Regelsystem zur Konvention, spricht man von einer Endogenisierung der Präferenzen.

Aspekte eines möglichen Auswegs - Glaubwürdigkeit und Bindung - herausgestellt, während im dritten Abschnitt Bindung vor dem Hintergrund der Verhandlungstheorie betont wird.

1.1 Eine Typologie sozialer Dilemmata

Ein soziales Dilemma sei in folgenden Fällen gegeben⁶:

- a) es existiert kein Gleichgewicht
- b) es existiert ein eindeutiges, aber ineffizientes Nash-Gleichgewicht ("Gefangenendilemma")
- c) es existieren mehrere nicht-pareto-vergleichbare Gleichgewichte und die Individuen wissen a priori nicht, welche Aktion ihr(e) Gegenspieler wählt bzw. wählen ("Chicken" bei Gleichgewichten in asymmetrischen bzw. "Battle of the Sexes" bei Gleichgewichten in symmetrischen Strategien)
- d) es existieren mehrere "pareto-ordenbare" Gleichgewichte und die Individuen kennen wiederum die Aktionen ihrer Gegenspieler nicht ("Assurance")⁷

Die verschiedenen Formen a) - d) eines sozialen Dilemmas werden nun in einem einheitlichen Rahmen anhand einer einfachen Spielstruktur verdeutlicht:

Zwei Spieler stehen vor der Wahl zwischen zwei möglichen Strategien: kooperieren (C) oder nicht kooperieren (D)⁸. Ihre Präferenzordnung (und dadurch unterscheiden sich die Situationen) wird durch die Höhe der Auszahlungen $U(C,C)$, $U(D,C)$, $U(C,D)$ und $U(D,D)$ ausgedrückt. Es handelt sich um ein 'one-shot-game' bei vollkommener Information, die Spieler treffen ihre Entscheidungen simultan. Es existiert also keine nächste Periode, die Spieler kennen die Präferenzordnung des Mitspielers und handeln gleichzeitig. Da beide Spieler die gleichen Handlungsmöglichkeiten besitzen, spricht man auch von einem symmetrischen Spiel. Die sich ergebenden Nutzenwerte (pay-offs) hängen sowohl von der eigenen Handlung als auch von der des Mitspielers ab, die Darstellung geschieht mittels einer 2x2-Matrix in der Normalform eines Spiels.

⁶ Anders motivierte Typologien finden sich bei Althammer/Buchholz (1995) oder bei Taylor (1987).

⁷ Außerdem wird ein bedeutender Spezialfall von d) diskutiert: es existiert ein pareto-superiores Gleichgewicht, aber die Spieler einigen sich nicht über die Aufteilung des Überschusses ("surplus-sharing-problem")

⁸ Die beiden Strategien können je nach Anwendungsfall unterschiedlich interpretiert werden. Eine der häufigsten Interpretationen ist die Beitragszahlung zur Finanzierung eines öffentlichen Gutes. "D" wird als Symbol für nicht-kooperierendes Verhalten verwendet, da dieses oftmals auch als "Defektion" bezeichnet wird.

Der erste Fall (a) eines sozialen Dilemmas ist das Fehlen eines Gleichgewichts. Innerhalb der gerade vorgestellten Spielstruktur kann man dabei zwei Versionen unterscheiden: in der ersten Version liegt ein Ungleichgewicht dann vor, wenn für den ersten Spieler "C" beste Antwort auf "C" und "D" beste Antwort auf "D" ist, während für Spieler 2 "C" beste Antwort auf "D" und "D" beste Antwort auf "C" ist. Die zweite Version beschreibt den umgekehrten Fall: hier ist für Spieler 1 "C" beste Antwort auf "D" und "D" beste Antwort auf "C", für Spieler 2 hingegen ist "C" beste Antwort auf "C" und "D" beste Antwort auf "D". Leicht verständlich werden diese Präferenzordnungen anhand einer Matrix-Darstellung:

	C	D
C	*4,3	1,4*
D	3,2*	*2,1

Abb.1a: Version 1 (kein GG)

	C	D
C	3,4*	*2,3
D	*4,1	1,2*

Abb.1b: Version 2 (kein GG)

Dabei wurden die jeweils besten Antworten mit * gekennzeichnet. Es existiert kein Aktionenpaar, das wechselseitig besten Antworten entspricht⁹. Die konkreten Zahlenwerte wurden zur Verdeutlichung gewählt.

Der zweite Fall (b) eines sozialen Dilemmas ist das **Kooperationsproblem** (Gefangenendilemma):

⁹ Dies gilt im angeführten Zahlenbeispiel allerdings nur, wenn man die Möglichkeit gemischter Strategien außer Acht läßt. Siehe zum Konzept gemischter Strategien auch die Bemerkungen bei der Entwicklung des entsprechenden Gleichgewichts im Chickenbeispiel.

	C	D
C	3,3	1,4*
D	*4,1	*2,2*

Abb.2 : Gefangenendilemma

Da für jeden der beteiligten Spieler D dominante Strategie¹⁰ ist, ist (D,D) einziges Nash-Gleichgewicht, obwohl (C,C) für beide einen höheren Nutzenwert liefern würde. (D,D) ist ineffizient. Ein Ausweg aus diesem Dilemma wäre nur möglich, falls beidseitige Kooperation durch eine glaubwürdige Androhung hinreichend starker Bestrafung im Nicht-Kooperationsfall erzwungen werden könnte bzw. falls die Möglichkeit oder das Interesse an Nicht-Kooperation nicht vorhanden wäre¹¹.

Im Gegensatz zum Gefangenendilemma sind die folgenden Typologien dadurch charakterisiert, daß mehrere Gleichgewichte existieren und damit die Frage nach der Auswahl eines der möglichen Gleichgewichte in den Vordergrund gerückt wird¹². In diesen Fällen spreche ich vom **Koordinationsproblem**¹³.

Falls sich die Präferenzen beider Spieler dadurch beschreiben lassen, daß es zwar für jeden am besten wäre, bei gegnerischer Kooperation selbst zu defektieren, aber im Falle der Nichtkooperation des Mitspielers eigene Kooperation vorgezogen wird, so liegt eine Präferenzstruktur vor, die als Chicken-Problem bezeichnet wird, die erste Version des Falls (c) eines sozialen Dilemmas:

¹⁰ Sowohl auf C als auch auf D ist D beste Antwort. Dies gilt für beide Spieler.

¹¹ Zum Problem der Kooperation aus sozialpsychologischer Sicht siehe Argyle (1991), aus anthropologischer und psychologischer Sicht den Sammelband von Hinde/Groebel (1991).

¹² Die Änderung der Spielstruktur vom one-shot-game zu einem wiederholten Spiel i.a. bei unendlichem Zeit-horizont kann zwar als *eine* mögliche Lösung eines Kooperationsproblems betrachtet werden, durch die unendlich große Menge möglicher Strategien führt dieser Weg aber geradewegs in ein Koordinationsproblem.

¹³ Eine Anwendung dieser terminologischen Unterscheidung auf makroökonomische Fragestellungen findet sich in Silvestre (1993). Als Grundlage dient dabei Marktmacht als Fähigkeit, Preise über Grenzkosten zu setzen. Daraus können sowohl Kooperations- (*ein* globales, aber ineffizientes Gleichgewicht) als auch Koordinationsprobleme (*mehrere* lokale Gleichgewichte) resultieren.

	C	D
C	1,1	*2,4*
D	*4,2*	1,1

Abb.3 : Chicken - Problem

Bei einer Chicken-Konstellation ist für beide Spieler C beste Antwort auf D und D beste Antwort auf C. Im Gegensatz zum Gefangenendilemma existieren jetzt zwei Nash-Gleichgewichte in reinen Strategien [(C,D) und (D,C)] sowie ein Gleichgewicht in gemischten Strategien ($p^* = q^* = 1/4$)¹⁴. Jedes dieser drei Gleichgewichte ist effizient. Welches der Gleichgewichte erreicht wird, hängt entscheidend davon ab, welchem Spieler es gelingt, den anderen glaubwürdig davon zu überzeugen, daß er auf keinen Fall bereit ist, kooperativ zu handeln, bzw. beide Spieler sich darauf verlassen können, daß der jeweils andere mit einer Wahrscheinlichkeit 1/4 kooperiert. Dann nämlich wird der Mitspieler allein (bzw. im angegebenen Verhältnis) "C" wählen und eines der drei Gleichgewichte wird erreicht.

Die zweite Variante des Falls (c) eines sozialen Dilemmas liegt dann vor, wenn für beide Spieler C beste Antwort auf C und D beste Antwort auf D ist. Im Gegensatz zum Chicken-Problem erhält man jetzt zwei Gleichgewichte in symmetrischen Strategien: (C,C) und (D,D)¹⁵. Wenn zusätzlich (C,C) und (D,D) nicht pareto-vergleichbar sind,

¹⁴ p (resp. q) gibt die Wahrscheinlichkeit an, mit der Spieler 1 (resp. Spieler 2) C spielt. Das Gleichgewicht in gemischten Strategien erhält man aus $Eu_1("C") = q + 2(1-q) = 4q + (1-q) = Eu_1("D")$. Daraus folgt $q^* = 1/4$. p^* erhält man aus $Eu_2("C") = p + 2(1-p) = 4p + (1-p)$. Ein Gleichgewicht in gemischten Strategien läßt sich auf zwei unterschiedliche Weisen interpretieren: zum einen verknüpft man damit die Vorstellung, daß jeder der Spieler im entsprechenden Verhältnis zwischen den möglichen Aktionen *lost*, also an einer Lotterie teilnimmt; zum anderen greift man auf einen Gedanken aus der evolutorischen Ökonomie zurück: im betrachteten Fall sei der Anteil der "C-Spieler" an der Gesamtpopulation aller Spieler gerade 1/4.

¹⁵ In Analogie zur Chickenstruktur existiert auch bei "Battle of the Sexes" als dritte Möglichkeit ein Gleichgewicht in gemischten Strategien.

spricht man vom sogenannten Battle of the Sexes¹⁶. Die Koordination erfordert hier die Abstimmung auf gleiches Verhalten¹⁷.

	C	D
C	*3,2*	1,1
D	1,1	*2,3*

Abb. 4: Battle of the Sexes

Beim Battle of the Sexes-Problem steht man im wesentlichen vor der gleichen Grundsituation wie bei Chicken. Der Unterschied liegt darin, daß jetzt der erste Spieler ein Interesse daran hat, seinem Mitspieler glaubhaft zu versichern, daß er auf alle Fälle kooperiert und es somit für den anderen am besten ist ebenfalls zu kooperieren. Für den zweiten Spieler ergibt sich hingegen die umgekehrte Situation: er wird versuchen Spieler 1 davon zu überzeugen, daß er auf alle Fälle *nicht* kooperieren wird und Spieler 1 mit der Erwiderung der Nichtkooperation besser fährt.¹⁸

Es findet also sowohl bei Chicken als auch bei Battle of the Sexes ein Wettbewerb um die glaubwürdige Signalisierung der Handlungsbeschränkung auf jeweils eine Aktion statt. Bei Chicken ist diese Aktion für beide Spieler 'Nichtkooperation', bei Battle of the Sexes für den ersten Spieler 'Kooperation', für den zweiten Spieler 'Nichtkooperation'.

Den vierten Fall d) eines sozialen Dilemmas liefert das Assurance-Problem. Durch eine Normalform-Darstellung wird es wie folgt charakterisiert:

¹⁶ Sind sie hingegen pareto-ordenbar, betrachtet man (siehe weiter unten) eine 'Assurance'-Struktur.

¹⁷ Sowohl bei Chicken als auch bei Battle of the Sexes sieht man sich mit einem Verteilungsproblem konfrontiert, während bei Assurance Effizienz im Vordergrund steht.

¹⁸ Experimentelle Überprüfungen von Battle of the Sexes - Spielen führten Cooper et. al. (1989, 1993) durch.

	C	D
C	*4,4*	1,2
D	2,1	*3,3*

Abb. 5: Assurance-Problem

Wieder gilt (wie bei Battle of the Sexes): C ist für beide Spieler beste Antwort auf C und D ist für beide Spieler beste Antwort auf D. Der entscheidende Unterschied liegt darin, daß (C,C) pareto-superior im Vergleich zu (D,D) ist. Worin das eigentliche Problem besteht, wird m.E. sehr gut deutlich, wenn man seine Aufmerksamkeit den Worten Amartya Sens schenkt, der als erster den Begriff "Assurance" geprägt hat:

"Expectations about other people's behavior must be brought in. If it is expected that the others will all do C¹⁹, then this one would prefer to do C also; otherwise he may do D. ... If everyone has implicit faith in everyone else doing the "right" thing, viz., C, then it will be in everyone's interest to do the right thing also. Then the outcome need not be Pareto-inferior. However, if each individual feels that the others are going to let him down, that is not to do C, then he too may do D rather than C, and the outcome will be Pareto-inferior. ... Given that each individual has complete assurance that the other will do C, there is no problem of compulsory enforcement. ... In this case assurance is sufficient and enforcement is unnecessary, and we shall refer to this case as that of the "assurance problem"."

Amartya Sen (1967), S.114f

Das pareto-superiore Gleichgewicht wird nur erreicht, wenn sich beide Spieler sicher sind, daß der Mitspieler ebenfalls kooperiert. Besteht nur eine geringe Unsicherheit, daß der Mitspieler nicht kooperiert, so kann (unter Berücksichtigung der pay-offs) der Erwartungswert $Eu(D)$ größer als $Eu(C)$ sein. Entscheidend ist also eine wechselseitige glaubwürdige Versicherung kooperativ zu handeln. Im zweiten Abschnitt wird unter anderem diskutiert, durch welche Ansätze versucht wurde, Glaubwürdigkeit bei Assurance zu begründen.

¹⁹ Sen verwendet im Original "B" für Kooperation (C) sowie "A" anstelle der Nichtkooperation (D).

Den in Fußnote 7 erwähnten Spezialfall von Assurance stellt das Surplus-Teilungsproblem dar. Kommt es bei der Existenz eines Surplus S nicht zu einer Einigung über die Aufteilung, kann als Endergebnis ein ineffizienter Zustand resultieren: es bleibt beim Ausgangszustand. Die Schwierigkeit besteht dabei in der Vielzahl möglicher Aufteilungen. Jede Aufteilung ist gegenüber einer Nichteinigung überlegen, bei der die Betroffenen mit leeren Händen den Verhandlungstisch verlassen. Bei geeigneter Interpretation der Aktionen C und D sowie entsprechender pay-off-Wahl läßt sich das Surplus-Teilungsproblem innerhalb des bisher verwendeten Rahmens diskutieren. Die Wahl von C sei dabei gleichzusetzen mit der Zustimmung zu irgendeiner Aufteilung ($g, 1-g$)²⁰, die Wahl von D bedeute die Ablehnung jeglicher Aufteilung. Die Blockade, sprich Wahl von D durch mindestens einen der beiden Spieler führt dazu, daß keiner der beiden Spieler einen positiven Anteil an S erhält. Vor diesem Hintergrund läßt sich dieses spezielle Assurance-Spiel einfach anhand unserer mittlerweile bekannten 2x2-Matrix beschreiben:

	C	D
C	$*g, 1-g*$	0,0
D	0,0	0,0

Abb. 6: Das Surplusteilungsproblem²¹

Obwohl aber beide Spieler danach trachten das Ergebnis der Nichteinigung zu vermeiden, ist das Erreichen von (C,C) nicht selbstverständlich²². Im Bewußtsein dieser Gefahr könnten die beiden Parteien nun auf die Idee kommen einen Schiedsrichter

²⁰ Wobei g den Anteil des Überschusses, den der erste Spieler erhält, beschreibt. Außerdem sei damit eine vollständige Aufteilung modelliert, d.h. der Überschuß S ist auf eins normiert: $g + 1 - g = S = 1$.

²¹ In diesem Fall gilt: C ist für beide Spieler beste Antwort auf C (wie im ursprünglichen Assurancespiel). Dagegen ist hier für beide Spieler sowohl D als auch C beste Antwort auf D. C ist also für beide Spieler schwach dominante Strategie.

²² Vgl. dazu z.B. die experimentellen Ergebnisse des sogenannten "Ultimatum-Spiels", nachzulesen bei Richard Thaler (1992), Kapitel 3. In Abschnitt 3 des letzten Kapitels verwende ich die Erkenntnisse aus diesem Spiel zur Motivation der Modellierung meines Fairness-Modells.

anzurufen, dessen Aufteilungsvorschlag von beiden akzeptiert wird. Damit ist die Bestimmung der Anteile an einen außenstehenden Dritten delegiert. Im Falle langandauernder Tarifkonflikte wird dieses Verfahren seit längerem erfolgreich praktiziert. Voraussetzung für das Funktionieren einer solchen Lösung ist natürlich die Akzeptanz der Person des Schlichters durch die Tarifparteien und dessen institutionell gesicherte Kompetenz der Durchsetzung des Kompromisses, mit anderen Worten: die Existenz glaubwürdiger Sanktionen bei Nichtkooperation. Ebenso wie eine natürliche Person kann man sich auch eine etablierte, von allen Gesellschaftsmitgliedern akzeptierte Regel als Schlichter vorstellen. Wie auch immer diese Regel ausgestaltet sein mag, sei es, daß der Ältere von beiden den doppelten Anteil erhält oder daß derjenige mit dem geringeren Vermögen $x\%$ des Surplus zugewiesen bekommt, einerlei - entscheidend ist nur die Befolgung einer, irgendeiner Regel. Dabei ergeben sich zwei Perspektiven: welche Regel sind die *Individuen* bereit anzuerkennen und welche Regel würde der *Staat* setzen? Die beiden Sichtweisen sind eng miteinander verbunden, da die Erfolgsaussichten staatlicher Politik sowohl von dem Machtpotential der Regierung als auch der "internen" Akzeptanz durch die Bürger beeinflußt werden. Militär, Polizei und/oder Geheimdienste als Formen staatlichen Zwangs sind eine Möglichkeit Verteilungsregeln durchzusetzen, die Übereinstimmung staatlich bestimmter Regeln mit in der Gesellschaft vorherrschenden Gerechtigkeitsvorstellungen eine andere²³. Glaubwürdigkeit wird im ersten Fall durch das Gewaltmonopol des Staates und durch die praktizierte Anwendung des Bestrafungsinstrumentariums im Nichtkooperationsfall gesichert, während im zweiten Fall die Existenz von Konventionen ursächlich für wechselseitiges kooperatives und damit koordiniertes Verhalten sorgt²⁴. Der Begriff der "Konvention" wird dabei im Sinne Robert Sugdens gebraucht²⁵. Was läge also näher, als ihn selbst zu zitieren?

²³ Ein etabliertes Rechtssystem enthält in diesem Sinn interpretierbare Regeln. Beispiele aus dem deutschen Recht für die Einbeziehung eines Aufteilungsprinzips in die Gestaltung einer Rechtsnorm sind z.B. im Einkommensteuerrecht, Familien- und Erbschaftsrecht oder auch im Gesellschaftsrecht (Gewinnanteilverhältnisse in Personengesellschaften: § 722 BGB für eine BGB-Gesellschaft oder § 121 Abs. 3 HGB für eine OHG) zu finden.

²⁴ Versuche, die evolutionäre Entstehung von Konventionen spieltheoretisch zu begründen, gehen zurück auf Axelrod (1984) und finden sich z.B. bei Sugden (1986) und Taylor (1987). Das dabei verwendete Gleichgewichtskonzept evolutionär stabiler Strategien geht zurück auf Maynard Smith (1982). Boyer/Orlean (1992) beschreiben eine Taxonomie des Konventionenwandels. Wärneryd (1990) untersucht die Bedeutung von Konventionen angewandt auf Sprache, Geld, Eigentum und auf die Unternehmung.

²⁵ Zur philosophischen Grundlegung siehe Lewis (1969).

"I shall define a convention as: any stable equilibrium in a game that has two or more stable equilibria. ... To say, that some strategy *I* is a stable equilibrium in some such game is to say the following: it is in each individual's interest to follow strategy *I* provided that everyone else, or almost everyone else, does the same. Thus a stable equilibrium may be understood as a self-enforcing rule. But not all self-enforcing rules are ones that would ordinarily be called conventions. A self-enforcing rule, I suggest, should be regarded as a convention if and only if we can conceive of some *different* rule that could also be self-enforcing, provided it once became established."

Robert Sugden (1986), S.32

Damit sollen die Ausführungen an dieser Stelle zu möglichen Auswegen aus einer Koordinationsproblematik einstweilen beendet werden. Der Vorgriff auf den dritten Abschnitt dieses Kapitels wurde deswegen gewählt, weil bei einem Assurancepiel intuitiv - und das gilt in diesem Fall mit am stärksten für Ökonomen - wohl die größten Schwierigkeiten beim bloßen Erkennen der Problematik zu vermuten sind. Denn warum, so wird häufig argumentiert, sollte sich eine pareto-superiore Lösung nicht durchsetzen? Ja, warum eigentlich nicht?²⁶

²⁶ Eine experimentelle Untersuchung dieser Frage ist zu finden bei: Cooper et. al. (1990) und van Huyck et. al. (1990).

In diesem Abschnitt wird eine akzentuierte Sicht der Theorie des Marktversagens präsentiert. Der Schwerpunkt wird auf den Zusammenhang zwischen externen Effekten und sozialen Dilemmata gelegt²⁷.

1.

Ein **externer Effekt** liegt vor, wenn die Entscheidung einer Person den Nutzen anderer Individuen beeinflusst. **Internalisierung** externer Effekte bedeutet die Aufhebung der Divergenz zwischen privaten und sozialen Konsequenzen. Dies kann auf zweierlei Arten geschehen: gelten sämtliche Annahmen einer Arrow-Debreu-Ökonomie, so sind pekuniäre Externalitäten die einzige Erscheinungsform der Externalitäten. Über die Marktpreise werden die Betroffenen vollständig kompensiert bzw. bestraft (kein Marktversagen). Gelten eine oder mehrere der Annahmen einer Arrow-Debreu-Ökonomie hingegen nicht, ist diese ausgleichende Funktion der Marktpreise gestört. Die sozialen Konsequenzen einer Aktivität finden sich nicht in gleicher Weise im Kalkül des Entscheiders, der **Markt versagt**.

2. - These -

Alle Gründe für Marktversagen lassen sich als externer Effekt begreifen.

²⁷ vgl. als Alternativen Laffont (1988) sowie die Aufsätze von Barr (1992): "Economic Theory and the Welfare State" und Inman (1987): "Markets, Governments, and the 'New' Political Economy". Barr unterscheidet zwischen "traditionellem Marktversagen" (unvollkommener Wettbewerb, externe Effekte, zunehmende Skalenerträge, öffentliche Güter, nicht geräumte Märkte, Einkommensexternalitäten und meritatorische Güter) einerseits und Informationsproblemen andererseits. Inman interpretiert sämtliche Marktversagensgründe als Gefangenendilemma: "In each instance of market failure .. agents were asked to reveal information about their benefits or costs from trades with no guarantee that that information would not be used against them. Without that guarantee, information is concealed, .. and the market institution fails." (Inman, S.672). Das kooperative Gleichgewicht wäre hier beidseitige wahrheitsgemäße Präferenzenthüllung.

Welche Spezialfälle eines externen Effekts führen zu Marktversagen?

3.1 Öffentliche Güter

Im Gegensatz zu privaten Gütern ist es bei **öffentlichen Gütern** nicht möglich, andere vom Konsum auszuschließen. Die Minimalfunktion des Staates, Sicherung der Eigentumsrechte, kann nicht vollständig greifen. Die Bereitstellung eines öffentlichen Guts wirkt sich deshalb auf alle Mitglieder der Gesellschaft aus.

Die natürliche Umwelt (Luft, Boden, Wasser), die Rechtsordnung oder auch das politische System eines Staates mögen als Beispiele für ein öffentliches Gut dienen.

Eine Internalisierung über den Markt würde voraussetzen, daß sich alle Nutzenänderungen, die durch die Bereitstellung eines öffentlichen Guts ausgelöst werden, in einem Marktpreis für das öffentliche Gut ausdrücken lassen. Die Samuelson-Bedingung vermittelt dies sehr anschaulich: sie fordert die Gleichheit der Grenzrate der Transformation und der *Summe* der Grenzzraten der Substitution. Der Markt versagt, da die einzelnen Akteure nicht bereit sind, für den Konsum des öffentlichen Guts zu zahlen, wenn sie bei Nichtzahlung von eben diesem Konsum nicht ausgeschlossen werden können (im Gegensatz zu einem privaten Gut).

Marktversagen bei öffentlichen Gütern wird in einer Situation asymmetrischer (Unmöglichkeit der Internalisierung durch private Verhandlungen) bzw. unvollkommener Information (Scheitern zentraler Enthüllungsmechanismen) zum unlösbaren Problem. Einer Kombination beider Problemkreise (Eigentum und Information) wird im nächsten Punkt eine separate Diskussion der nächsten Erscheinungsform eines externen Effekts, der asymmetrischen Information, vorangestellt.

3.2 Asymmetrische Information

Bei **asymmetrischer (privater) Information** ist der Nutzen des Uninformierten abhängig von den Angaben des Informierten bzgl. Qualität (adverse Selektion) oder Handlung (moralisches Risiko), die Handlung des Informierten ist abhängig von den Vorgaben des Uninformierten. Die wechselseitige Externalität wird in der Prinzipal-Agenten-Modellierung deutlich: eine Variable im Optimierungskalkül des Prinzipals (Uninformierten) wird zum Parameter im Kalkül des Agenten (Informierten).

Beispiele: Im ersten Fall dient der Gebrauchtwagenmarkt als Standardbeispiel²⁸. Der Käufer eines Wagens ist nur unzureichend über dessen Qualität informiert. Seine Kaufentscheidung und damit sein Nutzen sind abhängig von den Angaben des Verkäufers. Der zweite Fall wird häufig über den Arbeitsmarkt²⁹ illustriert. Ein Arbeitgeber kennt die Fähigkeiten eines potentiellen Arbeitnehmers nicht vollkommen. Die Entscheidung über die Einstellung des Arbeitnehmers ist abhängig von der vermuteten Qualifikation³⁰.

Der Ausgleich privater und sozialer Konsequenzen einer Handlung über Marktpreise setzt eine wahrheitsgemäße Informationsweitergabe an den Partner voraus. Der Anbieter eines Wagens schlechterer Qualität wie auch der potentielle Arbeitnehmer mit einer geringeren Qualifikation werden jedoch beide ein Interesse daran haben, sich als diejenigen mit höherer Qualität resp. Qualifikation auszugeben. Umgekehrt wird es das Bestreben der Anbieter besserer Qualität bzw. Qualifikation sein, den Nachfragern gerade diesen Umstand glaubwürdig zu signalisieren. Ein entsprechendes Signal (beispielsweise über Zertifikate oder Garantien) kann aber Kosten verursachen und seinerseits neue externe Effekte auslösen. Marktpreise verlieren die "reine" Funktion der Informationsübertragung aufgrund strategischer Überlegungen der Informierten.

Marktversagen läßt sich durch das Scheitern privater Verhandlungen begründen. Die Herausforderung durch Coase (1960) hat sicherlich dazu beigetragen, den Blick auf eine der Anforderungen seiner Effizienzthese, das Fehlen von **Transaktionskosten**, zu richten. Im Vergleich zur idealtypischen "Coase-Welt" führt die Existenz von Transaktionskosten³¹ dazu, daß private Verhandlungen nicht in der Lage sind, die Externalität zu internalisieren. Viele dieser Kosten können als Konsequenz asymmetrischer Informationen angesehen werden. Mittlerweile finden sich in der Literatur zahlreiche Modellierungen spezifischer Kosten. Ich werde einige von ihnen kurz charakterisieren und erläutern, warum sie durch asymmetrische Informationen verursacht sind.

²⁸ Die klassische Referenz ist Akerlof (1970).

²⁹ Diese Überlegungen finden sich erstmals bei Spence (1973).

³⁰ In dieser Form und Anwendung asymmetrischer Information finden sich auch die Ursprünge der Effizienzlohntheorie.

³¹ Die Bedeutung dieser Frage wird auch dadurch ersichtlich, daß sich die sogenannte "Transaktionskostenökonomie" als Teil der Institutionenökonomie in den letzten Jahren fest etabliert hat. Als Literatur möchte ich auf Williamson (1989) verweisen.

Ein Arbeitgeber kann nicht immer erkennen, ob ein Arbeitnehmer beabsichtigt, die Firma bald wieder zu verlassen. Ein höherer Lohn verringert für die Arbeitnehmer den Anreiz den Arbeitsplatz zu wechseln und vermeidet damit die Kosten der Neuausschreibung und des anschließenden Anlernens neu Beschäftigter.

Die Kontaktaufnahme mit Tauschpartnern kann Suchkosten nach sich ziehen. Darunter sind Stellenanzeigen der Unternehmen oder auch Bewerbungsschreiben der Arbeitssuchenden vorstellbar. Die Aufwendungen eines Akteurs können dabei die Erfolgswahrscheinlichkeit des anderen Akteurs beeinflussen.

Unter "menu costs" werden in der Literatur Kosten der Preisanpassung modelliert. Das Drucken neuer Preislisten, die Information der Kunden und Verkäufer durch Rundschreiben oder ähnlichem dienen als praktische Beispiele. Die Transaktion besteht hier in der Weitergabe von Information.

Notar- bzw. Rechtsanwaltskosten werden in Kauf genommen, da sie sich über das Rechtssystem als institutionalisierter Mechanismus zur Erzielung von Glaubwürdigkeit der Bestrafung im Falle eines Vertragsbruchs interpretieren lassen.

3.3 Zunehmende Skalenerträge

Zunehmende Skalenerträge der Produktionsfunktion bedeuten fallende Durchschnittskosten. Bei dieser Technologie wird ein Produzent bei einem Preis anbieten, der sich aus Grenzerlös = Grenzkosten ergibt³². Dies folgt, da er bei Preis = Grenzkosten (< Durchschnittskosten) Verluste machen würde und mehrere Anbieter sich (bei fallenden Durchschnittskosten) gegenseitig aus dem Markt verdrängen könnten. Es liegt ein "natürliches" Monopol vor. Die Möglichkeit der Preissetzung wirkt sich auf die anderen Marktteilnehmer aus: ein Argument der indirekten Nutzenfunktion eines Akteurs (der Preis des betreffenden Gutes) ist gleichzeitig Variable des Optimierungskalküls des "natürlichen" Monopolisten.

Zunehmende Skalenerträge können aber nicht nur bereits gegeben sein, sondern auch erst durch die Aktivitäten einzelner Akteure entstehen. So können z.B. die Forschungsanstrengungen eines Unternehmens dazu führen, daß sich der allgemeine Wissensstand erhöht. Läßt sich dieses erhöhte know-how nicht durch Patente vor der Nutzung durch andere schützen, bekommt es den Charakter eines öffentlichen Guts. Die Produktionsfunktion eines anderen Unternehmens kann durch die Forschungsausgaben des ersten Unternehmens so beeinflusst werden, daß sie zunehmende Skalenerträge aufweist. Dieses Ergebnis ist vor allem dann zu erwarten,

³² Dies gilt nicht für den Fall bestreitbarer Märkte.

wenn die Ergebnisse aus der Grundlagenforschung (entspricht einem hohen Fixkostenblock³³) anderen Unternehmen zur Verfügung stehen.

Das skizzierte Beispiel spiegelt eine bestimmte Form der Beeinflussung wider, die von Russell Cooper und Andrew John beschrieben wurde: "Strategic complementarities arise when the optimal strategy of an agent depends positively upon the strategies of the other agents."³⁴. Unter dem Punkt 6.2 wird der Zusammenhang zwischen strategischen Komplementaritäten und externen Effekten verdeutlicht.

Zunehmende Skalenerträge begründen innerhalb der "Neuen keynesianischen Makroökonomie (NKM)"³⁵ sehr häufig unvollkommenen Wettbewerb, oftmals in der Modellierung monopolistischer Konkurrenz. Damit wird Preiseinfluß der Akteure unterstellt, eine Externalität. Es resultieren Preis- und Lohnstarrheiten und daraus nicht-geräumte Märkte, es kommt zu Marktversagen.

4.

Ein **Kooperationsproblem** (Gefangenendilemma) ist dadurch gekennzeichnet, daß

- a) der kooperative Beitrag einer einzelnen Person für diese kostspieliger ist als der Anteil am Gesamtnutzenzuwachs, den sie durch ihren eigenen Beitrag ausgelöst hat und
- b) alle Beteiligten sich verbessern würden, wenn sich alle kooperativ verhielten.

Ein **Koordinationsproblem** ist durch multiple Gleichgewichte gekennzeichnet.

³³ Hier begegnen wir der Hypothese zunehmender Skalenerträge *der Erkenntnis*. Ist menschliches Lernen dadurch charakterisierbar, daß die Grenzerträge zunehmen, wenn wir ein grundlegendes Modell verstehen und variieren?

³⁴ Cooper/John (1988), S.441. Die Autoren verweisen als weitere Beispiele auf wechselseitige Abhängigkeit der Inputs eines Produktionsprozesses und Nachfrageexternalitäten.

³⁵ vgl. dazu vor allem Hargreaves Heap (1992a) und die dort zitierte Literatur; siehe aber auch Gordon (1990) und Abschnitt 4 in Silvestre (1993). Gordon bezeichnet preissetzendes Verhalten als die Essenz der NKM. In meiner Betrachtung externer Effekte habe ich davon abgesehen, unvollkommenen Wettbewerb als eigenen Punkt aufzuführen, da man nach der Begründung für diese Marktsituation zu fragen hat und damit bei einem der diskutierten Aspekte von Externalitäten landet.

5. - These -

Kooperations- oder Koordinationsprobleme entstehen aus einer nicht über den Markt internalisierten Externalität (Marktversagen) .

6.

Auf welche Weise resultieren Probleme der Kooperation und Koordination aus Marktversagen?

6.1 Öffentliche Güter und Asymmetrische Information

Das Ausgangsproblem bei der Bereitstellung öffentlicher Güter läßt sich als Gefangenendilemma beschreiben. Im Schulmodell eines einfachen 2-Personen Spiels gilt: Nichtkooperation ist dominante Strategie für beide Akteure. Betrachtet man die Annahmen des Modells, ist dieses Ergebnis alles andere als verwunderlich³⁶. Beide Individuen ziehen aus dem öffentlichen Gut positiven Nutzen (im Beispiel zwei Einheiten), welcher die individuellen Kosten (jeweils eine Einheit) übersteigt. Der größte Nutzenwert kann allerdings erreicht werden, wenn der Mitspieler allein für die Bereitstellung sorgt. Nur aufgrund der Nichtausschließbarkeit im Konsum wird Freifahrerverhalten überhaupt erst ermöglicht. Deswegen: Eine Beteiligung der Akteure an der Finanzierung des öffentlichen Gutes entsprechend der individuellen Nutzengewinne würde die positiven Externalitäten dieser Beitragszahlungen internalisieren.

Um dies zu erreichen, sind prinzipiell zwei Wege denkbar: Private Verhandlungen oder zentral organisierte Allokationsverfahren³⁷.

Verhandlungen a la Coase führen bei asymmetrischer Information zu ineffizienten Ergebnissen³⁸. Als Ausweg bietet sich ein staatlicher Mechanismus an, der die wahre Angabe der Präferenzen zur dominanten Strategie macht. Die Gruppe der Groves-Mechanismen erfüllt zwar diese Anforderung, schafft aber neue Probleme. Es ist nicht

³⁶ Es wird unter Verwendung der pay-offs aus dem Beispiel der Abb. 2 argumentiert.

³⁷ Die folgenden Ausführungen haben Haslbeck (1994) viel zu verdanken.

³⁸ vgl. dazu Buchholz/Haslbeck (1991/1992), Farrell (1987) oder Illing (1992).

möglich, das Aufkommen aus einer entsprechenden Steuer³⁹ anreizverträglich an die Spieler auszuschütten. Daraus läßt sich die Idee einer zweiten Gruppe staatlicher Anreize zur Präferenzenthüllung, der sogenannten AGV-Mechanismen⁴⁰, motivieren. Die Schwierigkeit, die aus den Groves-Mechanismen erwächst, wird als Anforderung bei der AGV-Gruppe berücksichtigt: ausgeglichener Staatshaushalt. Andererseits führen diese Mechanismen nicht dazu, daß die wahrheitsgemäße Offenbarung der Präferenzen dominante Strategie für beide Akteure ist. Stattdessen erhält man ein schwächeres Ergebnis. Die wahrheitsgemäße Offenbarung ist ein bayesianisches Gleichgewicht⁴¹. Daraus wiederum folgern zwei Problemkreise: multiple Gleichgewichte auf der einen Seite, Verteilung andererseits.

Die Gefahr multipler bayesianischer Gleichgewichte resultiert aus den für den jeweiligen Gleichgewichtstypus spezifischen Anforderungen an die wechselseitige Konsistenz⁴² von Erwartungen. Alle Spieler sind u.a. durch ihre subjektiven Wahrscheinlichkeiten gekennzeichnet. Ein Gleichgewicht wird durch die Menge der Strategien der Spieler beschrieben. Die einzelnen Aktionen innerhalb der jeweiligen Strategien werden in Abhängigkeit von den erwarteten Aktionen der Gegenspieler gewählt. Aus unterschiedlichen Vermutungssystemen der Spieler lassen sich deswegen i.a. verschiedene Gleichgewichte konstruieren.

³⁹ Das bekannteste Beispiel aus dieser Gruppe ist die Clarke-Steuer.

⁴⁰ Benannt nach d'Aspremont und Gerard-Varet (1979).

⁴¹ Bayesianische Gleichgewichte können als Erweiterung des Nash-Gedankens auf Spiele mit unvollständiger Information betrachtet werden. Von entscheidender Bedeutung in diesen Spielen ist die modellierte Vorstellung über die Bildung von subjektiven Wahrscheinlichkeiten. Eine Variante dieses Gleichgewichtskonzepts sind die sequentiellen Gleichgewichte, die im nächsten Kapitel verwendet werden.

Darstellungen dieser Refinements des Nash-Gleichgewichts finden sich mittlerweile in verschiedenen Lehrbüchern der Spieltheorie, z.B. in: Fudenberg/Tirole (1991) [formal], Kap.8; Rasmusen (1989), Kap.5 [mit Anwendungsbeispielen aus der Industrieökonomie]; Kreps (1990b), Abschnitt 12.7 [arbeitet mit Beispielen und macht dadurch die Technik verständlich] oder Binmore (1992a), Kap.11 [als erste Lektüre zu empfehlen].

⁴² Der Begriff der "Konsistenz" wird in der ökonomischen Literatur sehr häufig verwendet. Er ist der Logik entlehnt und entspricht im wesentlichen der Widerspruchslosigkeit. An dieser Stelle richtet sich die Forderung nach Konsistenz an die Entsprechung von vermuteten und tatsächlichen Aktionen resp. Vermutungen. Innerhalb der Bedingungen eines sequentiellen Gleichgewichts dagegen wird unter Konsistenz die Art und Weise des Aufdatierens von a-priori-Wahrscheinlichkeiten verstanden. Wie auch immer, eine auftretende Inkonsistenz innerhalb einer ökonomischen Argumentation dürfte zu den Todsünden schlechthin zählen.

Daneben stellt sich die Frage multipler Gleichgewichte auf einer zweiten Ebene. Die Ursache liegt in der strategischen (nicht-kooperativen) Spieltheorie selbst. Der Vorteil dieses Ansatzes, das Panoptikum möglicher Verhandlungsabläufe explizit zu modellieren, führt andererseits in direkter Linie nicht nur zu der Schwierigkeit der Eingrenzung der Lösungen innerhalb eines gegebenen Modells, sondern verstärkt auch die Probleme der Modellauswahl.

Selbst wenn die beiden Aspekte des Koordinationsproblem gelöst sind und beide Spieler die Strategien wählen, die zu einem Gleichgewicht mit wahrer Präferenzangabe führen, bleibt ein Verteilungsproblem. Dies gilt deswegen, da nach Myerson/Satterthwaite (1983) kein Mechanismus existiert, der gleichzeitig die Bedingungen des ausgeglichenen Staatsbudgets, der wahrheitsgemäßen Enthüllung privater Informationen und individueller Rationalität erfüllt. Die letzte Forderung besagt, daß es allen Spielern durch die Teilnahme am Spiel möglich ist, mindestens die Nutzenposition zu erreichen, die sie vor Beginn des Spiels bereits hatten. Eine Verletzung dieser Forderung würde dazu führen, daß die Zentralinstanz (mindestens) zwei pareto-nicht vergleichbare Situationen als Alternativen zur Wahl hätte.

6.2 Zunehmende Skalenerträge

Die first-best Lösung bei einem natürlichen Monopol wäre $\text{Preis} = \text{Grenzkosten}$ und Deckung der dadurch entstandenen Defizite durch eine lump-sum Steuer. Würde der Markt nicht versagen, müßte es möglich sein, in privaten Verhandlungen zwischen einem Produzenten mit der Technologie zunehmender Skalenerträge und einem oder mehreren Konsumenten zu folgendem Ergebnis zu kommen: Der Unternehmer nimmt die Produktion auf und stellt die Menge her, bei der $\text{Preis} = \text{Grenzkosten}$ erfüllt ist. Die Konsumenten finanzieren die dabei entstandenen Defizite.

Wieder ergeben sich gravierende Schwierigkeiten in einer Welt asymmetrischer Information. Der Produzent muß sich absichern, daß die Konsumenten nach Aufnahme der Produktion die höheren Preise auch tatsächlich bezahlen. Die Verträge, in denen sich die Konsumenten genau dazu verpflichten, müssen natürlich auch die Höhe der zu zahlenden Preise beinhalten. Diese Höhe ist aber abhängig von der Kostenfunktion des Produzenten. Die Information über die Kostensituation der Unternehmung wird im allgemeinen asymmetrisch verteilt sein. Der Versuch, in Verhandlungen eine wahre Angabe der Kosten zu erzielen, generiert jedoch Ineffizienzen. Zwischen den Konsumenten kann ein Kooperationsproblem bestehen. Dies gilt, wenn einerseits jeder Konsument ein Interesse an der Produktion des Gutes auch bei Zahlung des höheren

Preises besitzt, andererseits aber ein Konsum des Gutes zu einem Preis, der nur marginal über den Grenzkosten liegt, noch besser ist. Jeder einzelne Konsument hätte dann einen Anreiz, daß alle anderen für die Finanzierung des Defizits sorgen und er selbst anschließend zu einem geringeren Preis kauft⁴³.

Zur Verdeutlichung der Koordinationsfrage verwende ich das Beispiel strategischer Komplementaritäten. Der Kern des Arguments wird in einem einfachen formalen Gedankengang entwickelt.

2 Akteure maximieren ihren pay-off π , der sowohl von den eigenen als auch von den gegnerischen Anstrengungen e_1 und e_2 abhängt durch Wahl von e . Die pay-off-Funktionen seien gegeben durch:

$$(1) \quad \pi_2 = [ae_1^2 + (1-a)e_1]e_2 - \frac{1}{2}e_2^2$$

und

$$(2) \quad \pi_1 = [ae_2^2 + (1-a)e_2]e_1 - \frac{1}{2}e_1^2$$

Ein wechselseitiger externer Effekt liegt dann vor, wenn

$$(3a) \quad \frac{\partial \pi_1}{\partial e_2} \neq 0 \quad \text{und} \quad (3b) \quad \frac{\partial \pi_2}{\partial e_1} \neq 0$$

gelten.

Eine zusätzliche Form der Abhängigkeit besteht, wenn außerdem

$$(4a) \quad \frac{\partial^2 \pi_1}{\partial e_1 \partial e_2} > 0 \quad \text{und} \quad (4b) \quad \frac{\partial^2 \pi_2}{\partial e_2 \partial e_1} > 0$$

⁴³vgl. Inman (1987), S.659

gelten. In diesem Fall spricht man von einer wechselseitigen strategischen Komplementarität⁴⁴.

Aus den Bedingungen erster Ordnung lassen sich die Reaktionsfunktionen ermitteln:

$$(5) \quad R_2 = e_2(e_1) = ae_1^2 + (1-a)e_1$$

und

$$(6) \quad R_1 = e_1(e_2) = ae_2^2 + (1-a)e_2$$

Man erhält zwei pareto-vergleichbare Nash-Gleichgewichte N1 und N2:

$$(7) \quad N1 = (e_1, e_2) = (0, 0) \quad \text{mit} \quad (\pi_1, \pi_2) = (0, 0)$$

und

$$(8) \quad N2 = (1, 1) \quad \text{mit} \quad (\pi_1, \pi_2) = \left(\frac{1}{2}, \frac{1}{2}\right)$$

Da N2 gegenüber N1 pareto-superior ist, führt das Beispiel in ein Assuranceproblem.

Mit den Ausführungen in diesem Abschnitt 1.2 sollte die Bedeutung der Untersuchung sozialer Dilemmata in den Vordergrund gestellt werden. Da praktisch alle Gründe für Marktversagen in ein Problem der Kooperation und/oder Koordination münden, ist es schwer vorstellbar, daß die Frage Staat versus Markt, die Frage volkswirtschaftlicher Theorie ohne diesen Hintergrund sinnvoll diskutiert werden kann.

⁴⁴ Für $0 < a < 1$ sind (4a) und (4b) bei positiven e_i erfüllt. Eine Interpretation von (4b) aus der Umweltökonomie: π sei eine Umweltschadensfunktion, e_1 die Emission von SO_2 (saurer Regen) und e_2 die Emission von Schwermetallen.

2. Auswege ?

Nach der Vorstellung des Themas im Abschnitt 1.1, nach den Überlegungen zu seiner Bedeutung im Abschnitt 1.2, seien vor den beiden "Versuchen" hier und im folgenden (dritten) Abschnitt einige Bemerkungen zu möglichen Auswegen gestattet.

1.

Bei **Kooperationsproblemen** wird eine Lösung nur dann erreicht, wenn sich alle Spieler sicher sein können, daß alle anderen Spieler auf nicht-kooperatives Handeln verzichten. Diese Erwartungen können über entsprechende Präferenzen, eine Sanktionierung für den Fall der Nicht-Kooperation oder die Unmöglichkeit der Nicht-Kooperation begründet werden.

Auch bei **Koordinationsproblemen** werden Erwartungen zentral. Hier jedoch ist es notwendig, daß die Akteure ihre Erwartungen über das Verhalten des bzw. der anderen Akteure koordinieren: ein Spieler muß dem anderen glaubhaft versichern können, daß er die "richtige" Aktion (je nach Konstellation) wählt. Deshalb ist es bei Koordinationsproblemen notwendig, daß zwischen den Spielern eine Asymmetrie in den Präferenzen, in der Spielstruktur oder in den Restriktionen vorliegt. Denn dadurch wird die Ankündigung der Wahl einer bestimmten Aktion und damit des Spielens eines bestimmten Gleichgewichts glaubwürdig und die Abweichung vom symmetrischen Gleichgewicht begründet.

2.

Nun ist es notwendig sich darüber klar zu werden, wovon wir sprechen, wenn wir den Begriff der Glaubwürdigkeit verwenden, wenn wir ihn verwenden für die Akteure unserer spieltheoretischen Modelle. Glaubwürdigkeit bezieht sich auf eine Ankündigung zukünftigen Verhaltens, individuelle Zweckrationalität ist dabei beinhaltet⁴⁵. Daran schließt sich eine möglicherweise erstaunliche Beobachtung an: erst

⁴⁵ vgl. als einleitende Darstellung Hargreaves Heap (1992b), zu unterschiedlichen Rationalitätskonzeptionen in den Sozialwissenschaften Acham (1984), zum Unterschied zwischen substantieller und prozeduraler

mit Einbeziehung *irgendeiner* Theorie des Handelns werden zukünftige Aktionen, sofern von zusätzlicher Unsicherheit abgesehen wird, vorhersehbar. Glaubwürdigkeit entsteht somit durch den Übergang zur Sicherheit, ein Bereich, dessen Wesen der Sphäre des Glaubens entschwand. In anderen Worten: eine Ankündigung wird dann (z.B. im Sinne der Teilspielperfektheit) als glaubwürdig bezeichnet, falls sie innerhalb des spezifizierten Modells mit Sicherheit vorhersagbar wird.

Für weiter entwickelte Gleichgewichtskonzepte wird das Argument erweitert, nicht jedoch in seinem Kern verändert. Sicherheit betrifft dann nicht mehr nur die Vorhersagbarkeit zukünftigen Handelns, sondern auch die Vorhersagbarkeit zukünftiger Erwartungsbildung. Entscheidend ist also nicht die *Art* des Handelns oder der Erwartungsbildung, sondern die *Kenntnis dieser Art* beim Gegenspieler. Es ist in diesem Zusammenhang beispielsweise nicht relevant, ob ein Spieler bayesianisches Aufdatieren der Bildung von Wahrscheinlichkeiten an zukünftigen Entscheidungsknoten zugrunde legt oder einfach die alten Wahrscheinlichkeiten fortschreibt, bedeutsam ist nur, daß der andere Spieler Kenntnis von der Art des Aufdatierens besitzt. Dann entspricht Glaubwürdigkeit dem logisch-mathematischen Nachvollzug einer modellimmanenten Verhaltenshypothese.

Es ist von Bedeutung, noch einmal zu betonen: wenn die Kriterien der Entscheidungsfindung bekannt sind, wird die Prognose zukünftigen Verhaltens in einer sicheren Welt zur Nacherzählung in den Annahmen festgelegter Abläufe. Dies gilt auch für *unglaubwürdige* Ankündigungen. Der Unterschied besteht in der Übereinstimmung von Ankündigung des einen Spielers und logisch konsistenter Deduktion des anderen Spielers (glaubwürdig) einerseits sowie fehlender Übereinstimmung andererseits (unglaubwürdige Ankündigung). Meine Bemerkungen jedoch zielten nicht auf die *Entsprechung*, sondern auf die *Verbindung* von Glauben und Konsistenz.

Rationalität Simon (1976, 1978), als Überblicksartikel über "rational choice" Sugden (1991), zu den "Fallstricken" der Rationalität Elster (1987), zur Rationalität in der ökonomischen Modellbildung Vogt (1993), zu Rationalität und Spieltheorie Bicchieri (1992, 1993).

3. Ein kurzer Blick zurück

Die Vielfältigkeit der bisher vorliegenden Ansätze läßt unterschiedlich gesetzte Akzente erkennen. Da ich mich hier auf neoklassisch geprägte Ansätze konzentriere, unterteile ich nach den drei Komponenten eines neoklassischen Modells: Zielfunktion, Spielstruktur und Restriktion.

Innerhalb der **ersten Komponente** finden sich Vorstellungen über die Präferenzen der Akteure. Die häufigsten Ansätze finden sich zum Thema altruistischen Handelns⁴⁶. Die Statusabhängigkeit der Nutzenbewertung einzelner Güter rückt Frank (1987, 1988, 1989) in den Vordergrund⁴⁷.

Die Berücksichtigung dieser Motive eröffnet allerdings keinen *Ausweg* aus dem ursprünglich betrachteten Konflikt - der Konflikt liegt in solchen Fällen nicht oder nur in abgeschwächter Form vor. Aus diesem Grund erschien es mir auch nicht von primärem Interesse diese Richtung hier weiterzuverfolgen.

Die **zweite Komponente** ist sehr variantenreich, das Augenmerk wird hier auf verschiedene Schwerpunkte gelenkt. Beispielfhaft wären zu nennen: die Dauer des Kontakts⁴⁸, die Abfolge der zur Verfügung stehenden Aktionen⁴⁹, die Zahl der beteiligten Akteure⁵⁰ sowie die ihnen zugänglichen Informationen.

Die Bedeutung wiederholten Aufeinandertreffens nahm ich zum Anlaß für die Modellierung meines ersten "Versuchs". Dort steht - wie bereits erwähnt - die Bildung von Reputation auf der Grundlage der Bewertung früherer Aktionen im Vordergrund.

In der **dritten Komponente** schließlich unterscheide ich zwischen der Akzeptanz externer Restriktionen und der eigenen, freiwilligen Hinzufügung von Restriktionen.

⁴⁶ siehe z.B. Andreoni/Miller (1993), Andreoni (1990), Sugden (1982,1984), Simon (1993), Samuelson (1993) und Taylor (1987). Holländer (1990) verwendet soziale Zustimmung durch freiwillige Beitragsleistung als Variation des Altruismusgedankens. Eine philosophiegeschichtliche Perspektive findet sich bei Batson (1991), Kapitel 2.

⁴⁷ vgl. aber auch die Formulierung referenzabhängiger Nutzenfunktionen bei Kahneman/Tversky (1979) und Tversky/Kahneman (1991,1992).

⁴⁸ Dabei wird zwischen one-shot-play und wiederholten Spielen unterschieden, bei letzteren zwischen endlicher und unendlicher Wiederholung.

⁴⁹ Hier ist zwischen Modellen mit simultanen Offerten wie z.B. im Nash-Demand-Game versus Modellen mit alternierenden Offerten (als Standardmodell wäre hier das Rubinsteinispiel zu nennen) zu differenzieren.

⁵⁰ Man unterscheidet zwischen Zwei-Personen- und n-Personen-Spielen.

Die Beschränkungen können sich dabei auf die Handlungsmöglichkeiten, auf die Entscheidungskompetenz oder auf die Gestalt der Restriktion selbst beziehen.

Freiwillige Selbstbindung durch Verzicht auf eine Handlungsoption findet sich im zweiten "Versuch".

4.

Ein Ausweg aus einem sozialen Dilemma, sei es durch eine Modellierung im Rahmen der zweiten oder der dritten Komponente, ist an eine glaubwürdige Bindung der Akteure geknüpft.

3. Verhandlungen und Bindung - eine Akzentuierung

Ich beginne mit der Unterscheidung zwischen exogenen und endogenen Restriktionen. Innerhalb der zweiten Gruppe differenziere ich zwischen inter- und intrapersonellen Restriktionen (1.). Es folgen Bemerkungen zur Verhandlungstheorie und ihrer Beziehung zum Thema der Bindung (2.) Anschließend diskutiere ich die Nash-Verhandlungslösung als Standardmodell der axiomatischen Verhandlungstheorie (3.) sowie das Modell von Rubinstein als Beispiel der strategischen Verhandlungstheorie (4.).

1.

Die Geschichte beginnt im Grunde ganz einfach: Wir akzeptieren Einschränkungen unserer individuellen Freiheit oder sind sogar bereit, auf Spielräume, auf (Handlungs-) Möglichkeiten freiwillig zu verzichten.

1.1

Exogen gesetzte Restriktionen finden wir in allen Formen hierarchischer Beziehungen, in Staaten, in Unternehmungen, in religiösen Gruppen, in Familien und anderswo. Die Akzeptanz wird entweder durch Androhung einer Bestrafung bis hin zu totalitärer Konsequenz erzwungen oder durch innere Überzeugung getragen.

1.2

Endogen gesetzte Restriktionen beruhen auf der Einschätzung zukünftiger Ereignisse. In einer interpersonellen, interkulturellen oder internationalen Beziehung richtet sich diese Einschätzung an die Interpretation und Reaktion der anderen; in einem intrapersonellen Verhältnis an die zeitliche Dimension und Entwicklung des eigenen "Ichs"⁵¹. Mein heutiges "Ich" kann unter Umständen über mein zukünftiges "Ich" be-

⁵¹ Hier spricht die volkswirtschaftliche Literatur häufig von *Zeitinkonsistenz*.

stimmen, ein intrapersonelles Prinzipal-Agenten-Verhältnis, wenn man so will. Die klassischen Beispiele finden sich bei Homer und bei Goethe: Odysseus kann nur durch den Preis der Begrenzung seines zukünftigen Willens dem Gesang der Sirenen sowohl begegnen als auch entgehen⁵²; Faust schließt den Pakt mit Mephisto.

1.3

Mit Hilfe eines (wirtschafts)politischen Szenarios ist es möglich, eine Verbindung beider Aspekte endogener Bindung zu verdeutlichen. Die *intrapersonelle* Divergenz heutiger und zukünftiger Präferenzen von Politikern mag durch die Rahmenbedingung demokratischer Wahlzyklen motivierbar und dadurch für die Wähler erkennbar sein. Diese bestimmen nun wiederum durch ihr Wahlverhalten die Variation der Rahmenbedingung, von der die Präferenzen der Politiker abhängen. Um ein für ihre Zwecke günstiges Wahlergebnis zu erzielen, kann von den Politikern eine spezielle Form der Bindung gewählt werden, die eine glaubwürdige Ankündigung zukünftigen Verhaltens bedeutet und sich somit aus *interpersonellen* Gründen ergibt.

Als Beispiel der Bindung wird oftmals die Delegation der Entscheidungsbefugnis an eine andere Person, z.B. einen Notenbankpräsidenten, dem eine andere Position bei geldpolitischen Entscheidungen unterstellt wird, erwähnt.

Ausblick:

Im Versuch II untersuche ich ein Chickenproblem mit einem einfachen zweistufigen Modell, bei dem ein Akteur auf der ersten Stufe die Möglichkeit der Bindung besitzt. Dieses Modell wende ich zum einen auf ein Thema aus der Umweltökonomie, zum anderen auf das Motiv der Fairness an. Dazu und vor allem zu den Beweggründen, die mich veranlaßt haben, gerade diese Modellierung zu wählen, mehr zu Beginn von Versuch II.

2.

Zuvor möchte ich allerdings einen anderen Themenkreis beschreiten, der mit der Frage der Bindung in einem engen Zusammenhang steht, im weiteren Verlauf der gesamten

⁵² Erkennen wir nicht in der heutigen politischen Auseinandersetzung um die Einführung bzw. Ausweitung von Volksbegehren Anklänge an diese Vorstellungen?

Abhandlung aber nicht näher - z.B. in einem eigenen Versuch - behandelt wird. Innerhalb der Typologie habe ich zuletzt das Teilungsproblem als spezielle Form von Assurance vorgestellt. Dieses Problem wird in der volkswirtschaftlichen Diskussion mit der **Theorie der Verhandlungen** behandelt. Deshalb stelle ich zuerst die beiden Standardmodelle von John Nash (1953) und Ariel Rubinstein (1982) vor, anhand derer bereits die beiden Grundrichtungen charakterisiert sind. Kurze Anmerkungen zu der Beziehung zwischen Verhandlungstheorie und Gerechtigkeit begleiten diesen Teil. Daran anschließend kehre ich wieder zur Bindung zurück (die ich allerdings nie wirklich verlassen hatte).

Die Ursprünge der Verhandlungstheorie gehen zurück auf die Aufsätze von John Nash zu Beginn der 50er Jahre⁵³. Erinnern wir uns an die Grundproblematik eines Teilungsspiels: Effizienz wird nur durch die Existenz irgendeiner Teilungsregel gesichert, ihre spezielle Gestalt hingegen löst das inhärente Verteilungsproblem.

Eine Regel kann nun als exogen gegeben angenommen werden oder aber es kann der ambitioniertere Versuch unternommen werden, die Etablierung und Befolgung einer Regel (und damit wird die Regel zur Konvention) modellendogen zu begründen. Die erste Variante findet sich innerhalb der axiomatischen (auch als kooperativ bezeichneten), die zweite Variante innerhalb der strategischen (oder nicht-kooperativen) Verhandlungstheorie⁵⁴.

Während bei der strategischen Variante die Verhandlungsstruktur explizit modelliert wird, bestimmt man bei der axiomatischen Theorie das Ergebnis (die Anteile der beteiligten Individuen) über eine spezielle Funktion, die man deduktiv aus bestimmten Grundannahmen (Axiomen) ableitet. Bei der prinzipiell unendlich großen Anzahl möglicher Aufteilungen ist die Fähigkeit zur Einigung auf *eine* Aufteilung notwendige

⁵³ vgl. John Nash (1950, 1951, 1953)

⁵⁴ Schelling (1956) diskutierte bereits sehr früh einige der wesentlichen Kriterien der Lösung eines Verhandlungsspiels. Die Grundzüge beider Richtungen der Verhandlungstheorie (Nash und Rubinstein) finden sich mittlerweile in mehreren Lehrbüchern der Spieltheorie, so z.B. in Binmore (1992), Kap.5. Eine weitergehende Darstellung liefern Osborne/Rubinstein (1990) im ersten Teil ihres Buches. Zur axiomatischen Theorie siehe Kalai (1985) und Binmore (1987), zur Beziehung zwischen axiomatischer Verhandlungstheorie und "Social Choice" siehe Moulin (1989). Strategische Verhandlungstheorie wird von Sutton (1986), Binmore/Osborne/Rubinstein (1992) und Binmore et. al. (1992) beschrieben, strategische Ansätze bei privater Information stellen Kennan/Wilson (1993) vor. Eine Übersicht psychologischer Variablen, die sich auf das Ergebnis eines nicht-kooperativen Spiels auswirken können, ist bei Antonides (1991) zu finden.

Voraussetzung für das Vermeiden eines sozialen Dilemmas im Sinne der Nichtrealisierung von Surplusanteilen. Die kooperative Verhandlungstheorie versucht einen Ausweg gerade dadurch zu liefern, daß sich alle Spieler an den Axiomen des jeweiligen Modells orientieren und somit eine Offerte wählen (bzw. akzeptieren), die der aus den Axiomen generierten Zielfunktion entspricht. Aus der Tatsache, daß aus bestimmten Axiomen eine eindeutige Lösung erzeugt werden kann, kann aber noch lange nicht gefolgert werden, daß die Beteiligten dieser Lösung zustimmen würden. Die axiomatische Verhandlungstheorie liefert einen Ausweg aus einem sozialen Dilemma aber nur unter eben dieser Voraussetzung. Innerhalb einzelner Axiome können Gerechtigkeitsvorstellungen zum Ausdruck kommen⁵⁵. Die Befolgung einer daraus abgeleiteten axiomatischen Verhandlungslösung kann entweder über Zwang oder über die Interpretation eines Axioms als etablierte Konvention begründet werden⁵⁶.

Daraus ergibt sich eine Zwischenbemerkung zu den Begriffen: Die beiden Richtungen als kooperativ bzw. nicht-kooperativ zu bezeichnen, erscheint mir zumindest in manchen Fällen etwas unglücklich. Mit der diktatorischen Durchsetzung einer axiomatisch begründeten Teilungsregel läßt sich eine kooperative Vorstellung wohl nur schwer verbinden. Dies gilt im nicht-kooperativen Fall entsprechend beispielsweise im Fall der Modellierung einer Verhandlung zweier Altruisten.

Für die Lösung eines Teilungsproblems ist in welchem Fall auch immer der *bindende* Charakter der Teilungsregel entscheidend.

Diese Bindung kann in einer hierarchischen Situation durch Macht begründet werden. Macht entsteht dadurch, daß einer der Akteure Zugang zu Informationen, Technologie oder Handlungsoptionen besitzt, der dem anderen verwehrt bleibt. In dieser Überlegung finden wir wiederum die "Asymmetrie" als konstitutives Merkmal der Lösung eines Koordinationsproblems. Glaubwürdigkeit der Sanktionierung abweichenden Verhaltens, in diesem Fall von *einem* bestimmten Gleichgewicht abweichenden Verhaltens wird entweder durch die Beobachtung der Bestrafung oder durch die Beobachtung des zur Bestrafung zur Verfügung stehenden Instrumentariums

⁵⁵ Ob die Verhandlungstheorie überhaupt das geeignete Instrumentarium zur Erörterung von Gerechtigkeitsfragen sein kann, ist nicht unumstritten. Siehe dazu die Position von John Roemer (1986), auf die ich im Anschluß an die Vorstellung der Nash-Verhandlungslösung kurz eingehe.

⁵⁶ Zum Begriff der Konvention sei an die Bemerkungen in der Typologie erinnert.

erzielt. Also wieder: Glaubwürdigkeit durch Sicherheit oder durch eine *allen bekannte* Form der Bildung subjektiver Wahrscheinlichkeiten aus früheren Ereignissen.

Im zweiten Fall, der etablierten Konvention, führt wieder eine Asymmetrie, jetzt bezogen auf die wechselseitigen Erwartungen, zur Lösung des Koordinationsproblems. Die Konvention dient als zusätzliches Kriterium der Entscheidungsfindung. Zur Verdeutlichung: das klassische Beispiel der Straßenverkehrsordnung (StVO)⁵⁷. Alle Verkehrsteilnehmer fahren besser, wenn alle links oder aber alle rechts fahren, ohne Abstimmung dagegen ist es sicher am schlechtesten. Existiert nun bereits eine StVO, so dient sie als Konvention, als zusätzliches Kriterium der Selektion zwischen dem "Links-Gleichgewicht" und dem "Rechts-Gleichgewicht". Die Asymmetrie liegt hier in den etablierten Erwartungen bezüglich des Verhaltens aller anderen. Es ist für alle Beteiligten glaubwürdig, daß sich die jeweils anderen an diese Regel halten, da zum einen beobachtbar ist, daß abweichendes Verhalten in der Vergangenheit durch Bußgelder oder erlittene Unfälle bestraft wurde und zum anderen der Erwartungsnutzen für abweichendes Verhalten geringer als der Erwartungsnutzen bei regelkonformen Verhalten ist.

Nach diesen Bemerkungen ist es an der Zeit, die beiden Ansätze der Verhandlungstheorie genauer zu betrachten.

3.

3.1

Die **Nash-Verhandlungslösung (NVL)** ist das am häufigsten verwendete axiomatische Konzept. John Nash (1953) konnte zeigen, daß genau eine Funktion, das sogenannte Nash-Produkt, eindeutig aus bestimmten Axiomen folgt. Wendet man diese Funktion auf ein Aufteilungsproblem an, das durch die Zahl der Spieler, ihre Drohpunkte (d.h. ihre Auszahlungen bei Nichteinigung) sowie die aufzuteilende Menge charakterisiert ist, erhält man eine eindeutige Lösung. Die NVL lautet:

$$\max_{x^a, x^b} (U^a(x^a) - d^a)(U^b(x^b) - d^b) \quad \text{s.t. } x^a + x^b = X$$

⁵⁷ Die Wahl dieses Beispiel in Abweichung von der Teilungsproblematik möge verziehen werden.

Dabei ist X die zu verteilende Größe, x^a und x^b sind die Anteile der beiden Spieler mit den zugehörigen Nutzenwerten $U^i(x^i)$ bei Einigung, d^i die Nutzenwerte bei Nichteinigung für $i = a, b$. Das Verhandlungsproblem wird durch (X, d) charakterisiert, die Lösung des Problems sind die Anteile x^i . Die Handlungsanweisung lautet somit in Worten: "Wähle die Anteile x^a und x^b so, daß das Produkt aus den individuellen Differenzen zwischen Endergebnis und Resultat bei Verhandlungsabbruch maximal wird!" Das "Innenleben" dieser Regel sind die Axiome. Sie lauten: Invarianz bzgl. der Nutzenskalierung, Symmetrie, Unabhängigkeit bzgl. irrelevanter Alternativen und Effizienz. Eine kurze Diskussion wird ihren Inhalt verdeutlichen⁵⁸.

3.2

Das *Invarianzaxiom* besagt, daß die Lösung durch die Nutzenskalierung nicht beeinflusst sein sollte. Positive lineare Transformationen der Nutzenfunktion dürfen das Ergebnis nicht ändern. Deshalb wird eine Neumann-Morgenstern-Nutzenfunktion verwendet. Damit werden zwar keine interpersonellen Nutzenvergleiche möglich, jedoch Vergleiche intrapersoneller Bewertungen. Es lassen sich also keine Aussagen der Art "Joachim geht es doppelt so gut wie Peter" formulieren, wohl jedoch "für Joachim ist eine Flasche Alele doppelt so viele Zigaretten wert wie für Peter". Andere Vorstellungen über eine Aufteilung, wie z.B. "maximiere die Summe der Nutzengewinne" oder "gib dem Ärmere von beiden einen höheren Nutzengewinn", würden das Invarianzaxiom verletzen, sie benötigen die Möglichkeit interpersoneller Nutzenvergleiche.

Wenn die zu verteilende Menge *symmetrisch* bzgl. der 45°-Linie im x^a - x^b -Raum ist und d auf dieser Linie liegt, so muß die Lösung ebenfalls auf ihr liegen. Das bedeutet, daß innerhalb der NVL von Asymmetrien zwischen den Spielern, die nicht in (X, d) zum Ausdruck kommen, abstrahiert wird. Sind die Spieler in einem Verhandlungsproblem austauschbar, stellt das Symmetrieaxiom sicher, daß die Lösung beiden Spielern den gleichen Nutzen zuweist.

Die axiomatische Forderung nach *Unabhängigkeit bzgl. irrelevanter Alternativen* läßt sich einfach verdeutlichen: Falls die Lösung in einem Spiel mit Menge X in einem Verhandlungsspiel mit Menge $Z \subset X$ ebenso möglich ist, sollten beide Spiele zur gleichen Lösung führen. Der Haupteinwand lautet, daß die Lösung gerade nicht unabhän-

⁵⁸ Dabei werden die entsprechenden Abschnitte aus Elster (1989), S.54ff zugrundegelegt.

gig von der "technisch maximal möglichen" Position gesehen werden sollte⁵⁹. Kalai und Smorodinsky (1975) variierten die Nash-Verhandlungslösung gerade dadurch, daß sie das Unabhängigkeitsaxiom durch ein Monotonieaxiom ersetzten. Als Ergebnis ihres axiomatischen Systems resultiert die Handlungsanweisung: "Minimiere die maximale relative Konzession!", formal ausgedrückt:

$$\min_{x^i} (x_{\max}^i - x^i)(x_{\max}^i - d^i)!$$

wobei x_{\max}^i die "technisch maximal mögliche" Position bedeutet⁶⁰.

Effizienz schließlich verkörpert die plausible Forderung nach der Rationalität sämtlicher Akteure. Die Akzeptanz einer Aufteilung von beispielsweise 10% für die erste Teilnehmerin und 25% für die zweite ist schwer vorstellbar. Warum sollten Anteile vergeudet werden?

3.3

Es verbleiben drei Anmerkungen.

Zum einen: die NVL begünstigt den Spieler mit der geringeren Risikoaversion. Ein kurzes Beispiel von Ken Binmore (1992a, S.191-195) wird diesen Sachverhalt verdeutlichen.

Zwei Spieler mit Nutzenfunktionen $u_1(z) = z^\gamma$ und $u_2(z) = z^\delta$ verhandeln über die Aufteilung eines Dollars⁶¹, wobei $0 \leq \gamma, \delta \leq 1$ gilt. Der Anteil des ersten Spielers beträgt z , derjenige des zweiten Spielers dementsprechend $1-z$. Für $d_1 = d_2 = 0$ lautet das Optimierungsproblem:

⁵⁹ Der Unterschied in den Ansätzen von Nash und Kalai/Smorodinsky wird in einer graphischen Darstellung des Nutzenraums schnell deutlich, vgl. dazu die Abbildungen 2.2 (zu Nash) und 2.4 (zu Kalai/Smorodinsky) bei Osborne/Rubinstein auf den Seiten 16 bzw. 22.

⁶⁰ Für diejenigen unter den Lesern, die an einer ethischen Diskussion interessiert sind, sei darauf hingewiesen, daß David Gauthier (1986) zu zeigen versucht, daß die Lösung von Kalai/Smorodinsky von rationalen Individuen als Gerechtigkeitsprinzip anerkannt werden kann. Vgl. dazu aber auch die folgenden Hinweise auf die Aussagen John Roemers (1986).

⁶¹ Für $z > 0$ gilt: u_1 und u_2 sind streng monoton steigend und konkav. Damit sind beide Spieler risikoavers mit $u_i' > 0$ und $u_i'' < 0$.

$$\max z^{\gamma}(1-z)^{\delta} \text{ durch Wahl von } z !$$

Als Lösung ergibt sich:

$$(z^*, 1-z^*) = \left(\frac{\gamma}{\gamma + \delta}, \frac{\delta}{\gamma + \delta} \right)$$

Je risikoaverser ein Individuum also ist (d.h. je geringer δ, γ sind), desto geringer ist sein Anteil als Endergebnis der Nash-Verhandlungslösung⁶².

Zum zweiten: die NVL basiert auf einer symmetrischen Verteilung der *Grenznutzen-zugewinne*. Würden z.B. Caroline von Monaco und Antje Vollmer über die Aufteilung einer Erbschaft in Höhe von 1 Million DM verhandeln, so liegt der Grenznutzen von Frau Vollmer für eine zusätzliche DM sicher höher als bei Prinzessin Caroline. Dies wird durch die NVL dahingehend kompensiert, daß sich im Ergebnis für Prinzessin Caroline ein höherer DM-Betrag ergibt⁶³

Und schließlich: John Roemer. Er vertritt in einem Aufsatz, der im Oktober 1986 in *Ethics* erschien, die Ansicht, daß die axiomatische Verhandlungstheorie prinzipiell ungeeignet sei, um Fragen der Verteilungsgerechtigkeit zu behandeln: "I will argue in this paper against the application of bargaining theory to problems of distributive justice on grounds that the domain which bargaining theory takes - objects of the form $\{S, d\}$ ⁶⁴ - is informationally too impoverished to capture the important issues in distributive justice." (John Roemer (1986), S.89f). Unterstützung findet Roemer in einer empirischen Untersuchung von Yaari und Bar-Hillel (1984), in der gezeigt wurde, daß Menschen ethische Probleme unterschiedlich bewerten, je nachdem, ob es sich um Bedürfnisse, Wünsche oder Geschmack handelt. Unsere Vorstellungen hinsichtlich einer *gerechten* Verteilung hängen von der Situation ab⁶⁵. Betrachten wir dazu das Problem der Aufteilung von 12 Grapefruits und 12 Avocados auf zwei Personen, Jones und Smith. In einer ersten Variante gehen wir davon aus, daß Jones' Körper in der Lage ist, aus jedem Stück Grapefruit 100 mg Vitamin F zu gewinnen, aus den Avocados hingegen nichts. Bei Smith gilt: eine Grapefruit bringt ihm ebenso wie eine Avocado jeweils

⁶²Da gilt: $\frac{\partial z^*}{\partial \gamma} = \frac{\delta}{(\delta + \gamma)^2} > 0$.

⁶³Diese Argumentation geht zurück auf die Kritik der NVL durch Luce/Raiffa (1957, S. 129-132).

⁶⁴Fast überflüssig zu erwähnen: der Notation von John Roemer entspricht (X, d) im Text.

⁶⁵Das anschließende Beispiel findet sich bei Yaari/Bar-Hillel (1984) auf den Seiten 8ff.

50 mg Vitamin F. Wer soll nun wieviel vom jeweiligen Obst erhalten⁶⁶? In einer zweiten Version wurde dieses Problem etwas modifiziert. An die Stelle der spezifischen körperlichen Umsetzung in Vitamin F trat jetzt die unterschiedliche Zahlungsbereitschaft: Jones sei bereit, jede beliebige Menge an Grapefruit zu kaufen, vorausgesetzt, der Preis einer Grapefruit übersteigt \$1 nicht. Avocados liebt Jones jedoch nicht und ist infolgedessen nicht einmal bereit einen müden Cent für eine Avocado auszugeben. Smith dagegen schätzt die Avocado wie die Grapefruit und kauft beide, vorausgesetzt ihr jeweiliger Preis überschreitet die 50 cent Grenze nicht. Worin besteht nun die gerechte Aufteilung⁶⁷? - Die Ergebnisse aus der Untersuchung von Yaari/Bar-Hillel sprechen deutlich. Und: Informationen über Bedürfnisse, Wünsche oder Geschmack finden sich in (X,d) nicht.

Damit möchte ich allerdings diesen Gedankengang verlassen. Für eine weiterführende Auseinandersetzung sei der Aufsatz von John Roemer und die darin zitierte Literatur empfohlen. Hier ging es mir nur darum einige grundlegende Überlegungen zu vermitteln.

4.

Nach dieser Diskussion einzelner Gesichtspunkte der axiomatischen Modellierung komme ich zum alternativen Ansatz der nichtkooperativen Verhandlungstheorie. Hier steht der Verhandlungsprozeß im Vordergrund. Diese Perspektive liefert die axiomatische Theorie ja gerade nicht. Andererseits erhält man unglücklicherweise eine Unmenge von möglichen Modellvariationen, da zu jedem Verhandlungsablauf auch ein entsprechendes Modell formulierbar ist. Die Variationen beziehen sich auf die Zahl der Spieler (2-Personen oder n-Personen-Spiele), die Abfolge der Offerten (Simultanspiel oder sequentielles Spiel), die Anzahl der Begegnungen (one-shot-game oder wiederholtes Spiel mit endlichem oder unendlichem Zeithorizont), die zur Verfügung stehenden Aktionen sowie auf die Informationen über die Spielstruktur und das Verhalten der Mitspieler. Inwieweit eine axiomatische Lösung aus einem strategischen Ansatz entwickelbar ist, spielt eine entscheidende Rolle bei der Kritik an der Verwendung kooperativer Verhandlungstheorien. Der Vorwurf des 'ad hoc' wird

⁶⁶ Von insgesamt 163 jungen Frauen und Männern entschieden sich 82% für die Aufteilung (Jones: 8 Grapefruit, 0 Avocado, Smith: 4 G, 12 A). Als weitere Aufteilungen standen zur Wahl: (J:6-6, S:6-6) - 8%, (J:6-0, S:6-12) - 0%, (J:9-0, S:3-12) - 8% sowie (J:12-0, S:0-12) - 2%.

⁶⁷ Die Befragung brachte ein deutlich anderes Ergebnis: (J:6-6, S:6-6) - 9%, (J:6-0, S:6-12) - 4%, (J:8-0, S:4-12) - 28%, (J:9-0, S:3-12) - 24% sowie (J:12-0, S:0-12) - 35%.

durch eine Mikrofundierung natürlich entkräftet. Das am häufigsten verwendete axiomatische Konzept, die Nash-Verhandlungslösung ist durch die beiden Standardmodelle der nichtkooperativen Theorie von Rubinstein (1982) und Nash (1953) begründbar.

4.1

Das Modell von Rubinstein legt alternierende Offerten zugrunde. Spieler 1 beginnt in t_0 mit einem beliebigen Teilungsvorschlag $(z, 1-z)$ und Spieler 2 kann entweder akzeptieren oder ablehnen. Im ersten Fall ist das Spiel zu Ende, Spieler 1 erhält einen Anteil z , Spieler 2 entsprechend $1 - z$. Im zweiten Fall geht das Spiel in die zweite Runde t_1 . Spieler 2 kann nun seinerseits eine Gegenofferte formulieren, die Spieler 1 akzeptieren oder ablehnen kann. Dieser Ablauf wiederholt sich solange, bis einer der beiden Spieler der Offerte des anderen zustimmt. Zusätzlich wird im Rubinsteinmodell davon ausgegangen, daß beiden Spielern der Zeitpunkt, an dem die Einigung stattfindet, nicht gleichgültig ist. Deshalb werden zwei Diskontparameter δ_1 und δ_2 mit $0 < \delta_i < 1$ eingeführt. Eine Einigung in der jeweils nächsten Periode wird von den beiden Spielern entsprechend abdiskontiert. Als eindeutige (teilspielperfekte) Gleichgewichtslösung ergibt sich:

$$z^* = \frac{1 - \delta_2}{1 - \delta_1 \delta_2}$$

Die Höhe der Anteile wird somit durch die Diskontparameter bestimmt. Derjenige Spieler, der "ungeduldiger" ist, sprich den niedrigeren Diskontparameter besitzt, erhält den geringeren Anteil⁶⁸. Außerdem gibt es bei dieser Spielstruktur einen "first-mover-advantage"⁶⁹. Und: Gehen die Intervalle zwischen den einzelnen Perioden gegen Null, so konvergiert die Rubinstein-Lösung gegen die Nash-Verhandlungslösung⁷⁰.

4.2

John Nash erkannte sehr wohl die Notwendigkeit der Unterstützung seines axiomatischen Ansatzes durch eine explizite Verhandlungsstruktur. Deshalb stellte er selbst ein

⁶⁸Es gilt: $\frac{\partial z^*}{\partial \delta_1} = \frac{\delta_2(1 - \delta_2)}{(1 - \delta_1 \delta_2)^2} > 0$; $\frac{\partial z^*}{\partial \delta_2} = \frac{\delta_1 - 1}{(1 - \delta_1 \delta_2)^2} < 0$.

⁶⁹So ergibt sich z.B. für $\delta_1 = \delta_2 = 0,5$ im Gleichgewicht: $z^* = 0,66$.

⁷⁰Siehe dazu u.a. Binmore (1992), S.211f; ausführlichere Darstellungen finden sich bei Osborne/Rubinstein (1990), Kap.4 sowie bei Binmore/Rubinstein/Wolinsky (1986).

nicht-kooperatives Simultanspiel zur Diskussion: das Nash-Demand-Game⁷¹. Es gibt darin nur eine Periode ("one shot"), in der beide Spieler *gleichzeitig* einen Vorschlag über die Aufteilung der Ressource machen. Sind diese Vorschläge kompatibel, wird diese Lösung realisiert, ansonsten landen beide bei Nichteinigung. Da dieses Spiel eine sehr große Menge möglicher Gleichgewichte besitzt, schlug Nash als Refinement das "smoothed (bzw. perturbed) Demand-Game" vor. Die Idee bei diesem Ansatz besteht darin, bei den beiden Spielern Unsicherheit über die Größe des Verhandlungsgegenstands zu unterstellen. Läßt man diese Unsicherheit im Grenzwert gegen Null gehen, so resultiert die NVL.

⁷¹ Der Originalaufsatz ist Nash (1953). Unterschiedliche Präsentationen finden sich bei Binmore (1992a), Abschnitt 7.4 sowie bei Osborne/Rubinstein (1990), Abschnitt 4.3.

VERSUCH ÜBER DIE IDEE DER REPUTATION

(Zur Begründung von Kooperation in einem endlich wiederholten Spiel)

1. Reputation und wiederholte Spiele

Die Einleitung zu diesem Kapitel wurde zu einem großen Teil bereits vor neun Jahren geschrieben, nicht von mir und auch nicht in Kenntnis dieser Abhandlung. Dennoch dürfte es schwer möglich sein, die Motive, die sich mit meiner Modellierung von Reputation in Verbindung bringen lassen, deutlicher zu benennen, als es damals Robert Wilson tat. Deshalb trete ich zurück und übergebe ihm das Wort:

"In common usage, reputation is a characteristic or attribute ascribed to one person (firm, industry, etc) by another (e.g., "*A* has a reputation for courtesy"). Operationally, this is usually represented as a prediction about likely future behavior (e.g., "*A* is likely to be courteous"). It is, however, primarily an empirical statement (e.g., *A* has been observed in the past to be courteous"). Its predictive power depends on the supposition that past behavior is indicative of future behavior.

This semantic tangle can be unraveled by the application of game theory. In a sequential game, a player's strategy is a function that assigns the action to be taken in each situation (i.e., each possible information condition) in which he might make a choice. If the player has some private information (e.g., his preferences), then the choices of actions may depend on this information. In this case, others can interpret his past actions (or noisy observations that embody information about his past actions) as signals about what his private information might have been. More specifically, they can use Bayes' rule to infer from the history of his observed actions, and from a supposition about what his strategy is, a conditional probability assessment about what it is that he knows."

Robert Wilson (1985), S.27f

Das Grundmodell im folgenden Abschnitt wie auch die sich anschließenden Variationen tragen diese Ausführungen in sich. Ich untersuche, welche Bedingungen vorliegen müssen, damit eine Spielerin mit privater Information und "Nichtkooperation" als dominanter Strategie im stage-game, bei endlicher Wiederholung dennoch (?) kooperativ spielt⁷².

⁷² Die Theorie wiederholter Spiele fand Eingang in die Lehrbücher von: Friedman (1986), Kapitel 4; Rasmusen (1989), Kapitel 4 und 5; Kreps (1990b), Kapitel 14; Fudenberg/Tirole (1991), Kapitel 5 sowie Binmore (1992), Kapitel 8; siehe aber auch die Artikel von Sabourian (1989), Mertens (1990) sowie Pearce (1992) für wiederholte Spiele bei vollkommener Information.

Ein Überblick über Experimente, in denen Abweichungen vom "free-rider"-Verhalten beobachtet wurden, findet sich in Kapitel 2 von Thaler (1992).

Die erste Begründung von Kooperation in einem endlich wiederholten Spiel bei asymmetrischer Information lieferten Kreps/Milgrom/Roberts/Wilson (1982), als Weiterentwicklung sei Fudenberg/Levine (1989) erwähnt. Anwendungen in der Theorie der Unternehmung lassen sich nachlesen bei Kreps (1990a) und Miller (1992), zur Verwendung in der Industrieökonomie siehe Tirole (1988), Kapitel 6 und 9.

2. Das Grundmodell

Die Überprüfung von Kooperation in einem endlich wiederholten Spiel wird im anschließenden Modellrahmen, der so einfach wie möglich gehalten wird, präsentiert. Die notwendigen Modellbausteine ergeben sich aus der obigen Einführung: wir benötigen zwei Spielerinnen, von denen die eine über die Präferenzen der anderen nur unvollständig informiert ist. Insbesondere gehe Spielerin II mit einer Wahrscheinlichkeit p_1 davon aus, daß Spielerin I egoistische Präferenzen hat (dieser Typ sei mit I_E bezeichnet), mit der Gegenwahrscheinlichkeit $1 - p_1$ dagegen vermutet Spielerin II einer Altruistin (Typ I_A) gegenüberzustehen. Spielerin I dagegen kennt natürlich ihre eigenen Präferenzen, es wird also einseitig asymmetrische Information unterstellt. Beide Spielerinnen können in jeder Runde zwischen der Aktion "C" (entspricht "Kooperation") und der Aktion "D" (entspricht "Nicht-Kooperation" oder "Defektion") entscheiden.

Die Egoistin sei dadurch gekennzeichnet, daß für sie "D" dominante Strategie ist, die Altruistin ist durch "C" als dominante Strategie⁷³ und einen Altruismusparameter β , mit dem die pay-offs von Spielerin II in ihre Nutzenfunktion eingehen, charakterisiert. Für die Modellergebnisse ist nur die jeweils dominante Strategie notwendig, nicht jedoch die speziellere Modellierung mit β . Es lassen sich in diesem Sinne verschiedene Aspekte thematisieren, die Analyse ist nicht auf Altruismus beschränkt. Spielerin II besitzt "Koordinationspräferenzen", d.h. für sie gilt: beste Antwort auf "C" ist "C", beste Antwort auf "D" ist "D".

In einer ersten Version läuft das Spiel über zwei Perioden, in einer zweiten, allgemeineren dann über n Runden. Von einer Diskontierung wird abgesehen. In jeder der beiden Perioden wählen beide Akteure simultan zwischen "C" und "D". Diese Situation läßt sich anhand folgender Normalform charakterisieren:

⁷³ Deshalb sei $\beta > [y-x]/[a-d]$ wegen $x+\beta a > y+\beta d$ und $\beta > [w-z]/[b-c]$ wegen $z+\beta b > w+\beta c$, siehe Abbildung 1.

<div>II I E</div>		C	D
		C	D
C		x,a	z,b
D		y,d	w,c

"Egoistenspiel"

$\begin{array}{c} \text{II} \\ \text{I}_A \end{array}$		C	D
		C	D
C	$x+\beta a, a$	$z+\beta b, b$	
D	$y+\beta d, d$	$w+\beta c, c$	

"Altruistenspiel"

Abb.1: Normalform des Kooperationsmodells

Es gilt bzgl. der Präferenzen der Egoistin

$$(1) \quad y > x > w > z$$

sowie für Spielerin II

$$(2) \quad a > b > c > d$$

Es wird zwei Runden entweder das "Egoistenspiel" oder das "Altruistenspiel" gespielt. Spielerin I weiß, in welchem Spiel sie sich befindet, Spielerin II hingegen nicht. Im one-shot-"Egoistenspiel" resultiert (D,D) als Nash-Gleichgewicht - nicht überraschend in einer Gefangenendilemmasituation. Als Gleichgewicht im one-shot-Altruistenspiel dagegen erhält man (C,C).

Im weiteren wird jetzt untersucht, unter welchen Bedingungen sich im zweistufigen "Egoistenspiel" ein Gleichgewicht ergibt, in dem die Egoistin zumindest in der ersten Runde kooperiert. Als zusätzliche Komponente wird der Signalcharakter der Aktion der Egoistin in der ersten Runde berücksichtigt.

Die Zugfolge, die Spieldauer und die Wahrscheinlichkeiten, mit denen Spielerin II zu Beginn des Spiels damit rechnet, auf einen der beiden Typen von Spielerin I zu treffen, sind common knowledge. Desweiteren ist auch beiden Spielerinnen bekannt, wie sich eine Aktion von Spielerin I in der ersten Runde⁷⁴ auf das Vermutungssystem von

⁷⁴ Zur Notation. diese Aktion wird mit $a_{II}(t_1)$ für $i = \{ E ; A \}$ bezeichnet. Dies gilt für andere Aktionen analog.

Spielerin II, im folgenden mit μ_{II} bezeichnet, zu Beginn der zweiten Runde auswirkt: Spielerin II verändert ihre a-priori-Wahrscheinlichkeiten, falls sie neue Informationen mittels der Aktion von Spielerin I in der ersten Periode gewinnt, durch bayesianisches Aufdatieren und geht in t_2 von einer entsprechend korrigierten Wahrscheinlichkeit aus. Vor diesem Hintergrund stellt sich die Frage nach der optimalen Strategie für Spielerin I_E als klassischer trade-off in einem Signalspiel: wenn sie in t_1 "C" wählt, verzichtet sie auf einen höheren pay-off heute. Dies ist nur dann sinnvoll, wenn es ihr gelingt, durch "C" die Vermutungen ihrer Mitspielerin so zu beeinflussen, daß Spielerin II im Glauben einer Altruistin gegenüberzustehen, ihrerseits in der zweiten Runde "C" spielt.

Zuvor jedoch noch ein Wort zum verwendeten *Gleichgewichtskonzept*: am einfachsten wäre sicherlich das Kriterium der Teilspielperfektheit. Dazu ist jedoch eine eindeutige Identifikation sämtlicher Teilspiele notwendig. Zentral an dieser Stelle ist aber gerade die Unsicherheit von Spielerin II über den Typus von Spielerin I. Spielerin II weiß also nicht, in welchem Teilspiel sie sich befindet. Deswegen kommt das sequentielle Gleichgewichtskonzept⁷⁵ zum Tragen. Es verkörpert die Übertragung des Gedankens der Teilspielperfektheit auf Spiele mit unvollkommener Information. Ein Gleichgewicht in diesem Sinne besteht aus zwei Komponenten: den Strategien der beiden Spielerinnen (dem Strategieprofil) und - zusätzlich - dem Vermutungssystem der Spielerin mit unvollkommener Information, hier Spielerin II. Das Vermutungssystem beschreibt die Wahrscheinlichkeiten, die Spielerin II den Knoten in sämtlichen Informationsmengen des gesamten Spiels zuweist. Für den Nachweis eines sequentiellen Gleichgewichts sind zwei Schritte notwendig: zum einen ist zu zeigen, daß die Wahrscheinlichkeiten in späteren Perioden in konsistenter Weise aus den bisherigen Aktionen und früheren Wahrscheinlichkeiten durch bayesianisches Aufdatieren entwickelt werden können; zum anderen muß die Handlungsweise beider Spielerinnen dem Kriterium der sequentiellen Rationalität entsprechen.

Der Grundgedanke läßt sich wohl am einfachsten verdeutlichen, wenn man ein Szenario wählt, in dem die Anreizbedingungen für Kooperation in der ersten Runde für eine Egoistin erfüllt und der zweiten Spielerin bekannt sind. Wir beginnen damit, die Geschichte aus der Sicht von Spielerin II nachzuvollziehen.

⁷⁵ Dieses Konzept wurde von David Kreps und Robert Wilson (1982) eingeführt.

2.1 Ein Poolgleichgewicht⁷⁶

Welche pay-offs in der ersten bzw. in der zweiten Runde hat II zu erwarten, falls sie ihrerseits jeweils zwischen "C" und "D" zu entscheiden hat? Dies hängt natürlich davon ab, auf wen sie trifft und mit welcher Aktion der Gegenspielerin sie rechnet. In t_1 und auch in t_2 wird eine Altruistin ihre dominante Strategie "C" wählen, bei einer Egoistin jedoch ist die Vorhersage für Spielerin II nicht so einfach.

Die Aktion einer Egoistin in t_2 wird immer "D" sein, da danach das Spiel beendet ist und dieser Aktion somit keinerlei Signalcharakter mehr zugeschrieben werden kann. "D" ist ja annahmegemäß dominante Strategie im stage-game. Welche Aktion der Egoistin wird II aber in der ersten Runde erwarten? Um diese Frage klären zu können, ist es notwendig, die pay-offs der Egoistin bei der Wahl der Strategie $\{C, D\}$ versus den pay-offs bei Wahl der Strategie $\{D, D\}$ zu ermitteln und zu vergleichen. Eine Abweichung von "D" in t_1 kann nur durch die Wirkung dieser Wahl auf die Vermutungen von II zu Beginn der nächsten Runde begründet werden. Durch "D" gibt sich die Egoistin zu erkennen, da eine Altruistin ja niemals defektieren würde. Die einzige Möglichkeit, II keine Sicherheit über den Typus der Gegenspielerin zu verschaffen, besteht für die Egoistin folglich in der Wahl von "C". Dies bedeutet aber eine Nutzeneinbuße in t_1 . Damit sich $a_{IE}(t_1) = "C"$ trotzdem lohnt, muß ein höherer Ertrag in der nächsten Runde diesen Verlust kompensieren. Wenn eine Egoistin davon ausgeht, daß II in t_1 "C" spielt (was anschließend begründet wird), liefern die beiden zur Wahl stehenden Strategien folgende pay-offs:

Φ_{IE}	$\{D, D\}$	$\{C, D\}$
u_{IE}	$y + w$	$x + y$

⁷⁶ In einem Poolgleichgewicht lernt die uninformierte Spielerin nichts aus den Aktionen der Gegenspielerin in der vorausgegangenen Periode, da beide Typen (die Egoistin wie auch die Altruistin) die gleiche Aktion wählen. Die Wahrscheinlichkeiten zu Beginn von t_2 entsprechen daher den a-priori-Wahrscheinlichkeiten $(p_1, 1-p_1)$. Die Revision ursprünglicher Wahrscheinlichkeiten durch bayesianisches Aufdatieren (im Rahmen eines sequentiellen Gleichgewichts) wird in späteren Modellversionen verdeutlicht.

Für $x > w$ folgt: $\Phi_{IE}^* = \{ C, D \}$ und damit das Ergebnis der Kooperation einer Egoistin in t_1 .

Die pay-offs in t_1 ergeben sich unter der erwähnten Behauptung $a_{II}(t_1) = "C"$. Bei obigen pay-offs ist außerdem berücksichtigt, daß Spielerin II in der zweiten Runde "C" wählt, falls Spielerin I in der ersten Runde ebenfalls "C" gespielt hat. War jedoch $a_I(t_1) = "D"$, so wählt Spielerin II in t_2 "D". Dies gilt es natürlich im folgenden zu zeigen.

Um die pay-offs in der zweiten Runde zu ermitteln, ist es für die Egoistin notwendig zu wissen, welche Aktion II in der zweiten Runde wählt. Dies ist wiederum davon abhängig, mit wem II glaubt zu spielen. Die Vermutung $(p_2, 1-p_2)$ ist aber abhängig von der Aktion der Egoistin in der ersten Runde. Deswegen muß man zurück in die Bestimmung des Erwartungsnutzens der Spielerin II in t_1 . Dieser ergibt sich als:

$$(3) \quad Eu_{II}(t_1) = \begin{cases} p_1 a + (1-p_1)a = a & \text{if } a_{II}(t_1) = "C" \\ p_1 b + (1-p_1)b = b & \text{if } a_{II}(t_1) = "D" \end{cases}$$

In den $(1-p_1)$ -Fällen, in denen II auf eine Altruistin trifft, rechnet sie mit "C" als Aktion der Mitspielerin und erhält a bzw. b bei eigener Kooperation bzw. Defektion. Bei einer Egoistin vermutet sie ebenfalls "C", da ihr $x > w$ bekannt ist⁷⁷. Wieder liefert eine Kooperation den höheren Nutzen a. Daraus folgt die oben erwähnte Behauptung: in der ersten Runde spielt II "C". Für die Entscheidung in Runde 2 stellt sich zuerst die Frage nach den subjektiven Wahrscheinlichkeiten, mit denen II jetzt vermutet, in ein Egoisten- oder in ein Altruistenspiel zu geraten. Da sie aus der Aktion $a_{II}(t_1) = \{C\}$ ihrer Gegenspielerin keine zusätzlichen Informationen gewinnen konnte, verwendet sie wiederum die a-priori-Wahrscheinlichkeiten $(p_1, 1-p_1)$. Als Erwartungsnutzen ergibt sich:

$$(4) \quad Eu_{II}(t_2) = \begin{cases} p_1 d + (1-p_1)a & \text{if } a_{II}(t_2) = "C" \\ p_1 c + (1-p_1)b & \text{if } a_{II}(t_2) = "D" \end{cases}$$

Im Gegensatz zu t_1 wird eine Egoistin in t_2 "D" spielen. Deshalb ergibt sich als pay-off für die Spielerin II in t_2 , falls sie auf eine Egoistin trifft, d bei eigener Wahl von "C" sowie c bei eigener Wahl von "D". Spielerin II bevorzugt in t_2 somit "C" gegenüber "D", falls

⁷⁷Die Überlegungen von Spielerin II in den Augen der Egoistin und umgekehrt verlaufen simultan.

$$(5) \quad p_1 < \frac{a-b}{a-b+c-d}$$

gilt. Nur wenn die a-priori-Wahrscheinlichkeit auf eine Egoistin zu treffen nicht zu groß ist, wird II in der zweiten Runde kooperieren und die "nicht entlarvte" Egoistin damit ihr Ziel erreichen. Wie entscheidend die Täuschung ist, wird sehr schnell deutlich, wenn man sich überlegt, was passieren würde, wenn die Egoistin in der ersten Runde "D" spielen würde. Daraufhin kann II erkennen, mit wem sie es zu tun hat und ihre Wahrscheinlichkeiten (p_2 , $1-p_2$) entsprechend neu bestimmen: sie setzt in diesem Fall p_2 (die Wahrscheinlichkeit auf eine Egoistin zu treffen) gleich eins. Aus (4) folgt dann (für $c > d$): II wählt in t_2 "D".

Mit diesen Überlegungen läßt sich die Wahl von $a_{II}(t_2) = "C"$ und damit die oben beschriebenen pay-offs für die beiden betrachteten Strategien der Egoistin { C , D } und { D , D } begründen. Das sequentielle Gleichgewicht lautet:

Das Vermutungssystem μ_{II} von Spielerin II ist gegeben durch:

$$(p_1, 1-p_1) \text{ mit } p_1 < \frac{a-b}{a-b+c-d} < 1$$

$$(p_2, 1-p_2) = \begin{cases} (p_1, 1-p_1) & \text{if } a_I(t_1) = "C" \\ (1, 0) & \text{if } a_I(t_1) = "D" \end{cases}$$

(6)

Das Strategieprofil $\{\Phi_i^*\}$ ist gegeben durch:

$$\Phi_{IE}^* = \{ a_{IE}(t_1); a_{IE}(t_2) \} = \{ C; D \}$$

$$\Phi_{II}^* = \{ a_{II}(t_1); a_{II}(t_2) \} = \{ C; C \}$$

$$\Phi_{IA}^* = \{ a_{IA}(t_1); a_{IA}(t_2) \} = \{ C; C \}$$

2.2 Ein Separationsgleichgewicht

In diesem Unterabschnitt wird ein Gleichgewicht vorgestellt, in dem Spielerin II aus der Aktion ihrer Gegenspielerin in der ersten Runde neue Informationen gewinnt und ihre ursprünglichen Vermutungen daraufhin revidiert. Für diesen Fall ist eine neue Annahme hinsichtlich der Vermutungen von Spielerin II über den Anreiz für die Egoistin in t_1 zu kooperieren notwendig. Dazu wird eine neue Variable α eingeführt. Dies wird an der entsprechenden Stelle näher erläutert. Wir beginnen jetzt aber wieder unseren Weg durch die Überlegungen der Spielerinnen.

Am Anfang von Periode 1 (t_1) entscheiden beide Spielerinnen simultan über die Wahl zwischen "C" und "D". Der Nutzen aus der jeweiligen Aktion ergibt sich für Spielerin II in Abhängigkeit vom Typus ihrer Gegenspielerin. Desweiteren muß sich Spielerin II überlegen, welche Aktion eine Egoistin bzw. eine Altruistin wählen wird, um die Erwartungsnutzen der beiden möglichen Aktionen kalkulieren zu können. Dabei ergibt sich⁷⁸:

$$(7) \quad Eu_{II}(t_1) = \begin{cases} p_1 d + (1-p_1)a & \text{if } a_{II}(t_1) = "C" \\ p_1 c + (1-p_1)b & \text{if } a_{II}(t_1) = "D" \end{cases}$$

Zur Erinnerung: mit p_1 glaubt Spielerin II auf eine Egoistin, mit $1 - p_1$ hingegen auf eine Altruistin zu treffen. Eine Altruistin wurde gerade so charakterisiert, daß "C" für sie dominante Strategie ist. Daraus ergeben sich die entsprechenden pay-offs b bzw. a aus der Normalform lt. Abbildung 1 bei Wahl von "D" bzw. "C" durch Spielerin II.

Die Bestimmung der pay-offs im Egoistenfall gestaltet sich etwas schwieriger. Um diese Frage klären zu können, muß sich Spielerin II in die Position von Spielerin I_E versetzen. Die Egoistin erhält in t_1 (wenn sie davon ausgeht, daß Spielerin II in t_1 ihrerseits "D" wählen wird, was wiederum anschließend erläutert wird):

$$(8) \quad u_{IE}(t_1) = \begin{cases} z & \text{if } a_I(t_1) = "C" \\ w & \text{if } a_I(t_1) = "D" \end{cases}$$

⁷⁸Im Unterschied zu (3) vermutet II jetzt, daß eine Egoistin in der ersten Runde "D" wählt. Dies wird über geringere a-prioris ($p_1, 1-p_1$) begründet. (5) sei hier also gerade nicht erfüllt.

Bei Wahl von "C" würde sie somit $w - z$ verlieren. Um die optimale Strategie für Spielerin I_E ermitteln zu können, muß man sich zusätzlich die Signalfunktion von $a_I(t_1)$ vor Augen halten. Durch die Wahl von "C" bringt sie Spielerin II dazu in t_2 ebenfalls "C" zu wählen, während $a_I(t_1) = "D"$ eine Defektion von Spielerin II in t_2 zur Folge hätte (auch diese Behauptung wird unten begründet). Die Egoistin wählt in t_2 auf alle Fälle "D", da in der letzten Runde keine Signalfunktion mehr besteht. Bei erfolgreicher Signalisierung ergäbe sich deshalb in Runde 2 ein "Gewinn" in Höhe $y - w$, der den "Verlust" $w - z$ in der ersten Periode überkompensieren muß, damit eine Signalstrategie gewählt wird. Die optimale Strategie Φ_{IE}^* wird bei einem Vergleich noch einmal erkennbar:

Φ_{IE}	$\{ D , D \}$	$\{ C , D \}$
u_{IE}	$w + w$	$z + y$

Da $a_{IE}(t_2) = "D"$ dominante Strategie ist, kann die Wahl nur zwischen $\{C,D\}$ und $\{D,D\}$ liegen. Damit $\{C,D\} \succ \{D,D\}$ gilt, muß die sogenannte Signal-Anreizbedingung erfüllt sein, d.h. $a_{IE}(t_1) = "C"$ ist optimal, iff

$$(9) \quad y - w > w - z$$

gilt. Um Kooperation begründen zu können, dürfen die Kosten einer Signalisierung also nicht zu hoch sein. Diese Relation sei im weiteren gerade erfüllt.

Im Vergleich zum vorhergehenden Abschnitt läßt sich dieses Ergebnis einfach interpretieren: Falls der Anteil der Egoistinnen in der Population (p_1) den kritischen Wert $[a-b] / [a-b+c-d]$ überschreitet, muß die zusätzliche Restriktion (9) hinsichtlich der pay-offs der Egoistinnen erfüllt sein, damit Kooperation gezeigt werden kann.

Diese Überlegungen der Egoistin vollzieht Spielerin II nach, wenn sie die pay-offs in (7) ermittelt. Der oben skizzierte *trade-off* (9) wird entscheidend für die Erwartungshaltung von Spielerin II. (7) ergibt sich, wenn Spielerin II damit rechnet, daß eine Egoistin "D" spielen wird. Deshalb ist an dieser Stelle die eingangs erwähnte

zusätzliche Annahme notwendig: Spielerin II sei im Glauben, daß $y - w < w - z$ gilt, während in Wirklichkeit genau die umgekehrte Größenrelation zutrifft⁷⁹. Ein Vergleich der beiden Erwartungswerte aus (7) zeigt, daß "D" \succ "C", wenn

$$(10) \quad p_1 > \frac{a - b}{a - b + c - d}$$

gilt. Die entscheidenden Modellergebnisse lassen sich bei niedrigem p_1 für $a_{II}(t_1) = \text{"C"}$ (siehe den vorausgegangenen Abschnitt bei Gültigkeit von (5)) als auch für $a_{II}(t_1) = \text{"D"}$ zeigen. In diesem Abschnitt wird der Fall diskutiert, in dem p_1 hinreichend groß ist, sodaß (10) erfüllt ist.

Spielerin II muß ebenso ihren Erwartungsnutzen in Periode 2 ermitteln. Dabei wird der Dreh- und Angelpunkt dieses Ansatzes deutlich: die Abhängigkeit der Wahrscheinlichkeiten zu Beginn der zweiten Runde von den Aktionen der Spielerin I in Runde 1. Deswegen ist hier eine Fallunterscheidung vorzunehmen:

Fall 1: $a_I(t_1) = \text{"C"}$

Die entscheidende Frage ist, welche Schlüsse Spielerin II zieht, wenn sie "Kooperation" ihrer Gegenspielerin in der ersten Runde beobachtet. Eine Altruistin hätte in Runde 1 auf alle Fälle "C" gespielt, da für diese "C" dominante Strategie ist. Bei einer Egoistin hängt alles davon ab, ob "C" von Spielerin II als Signal für eine Altruistin verstanden wird und diese somit in Runde 2 ebenfalls "C" wählt. Dieser Konstellation ist sich Spielerin II bewußt. Wenn sie davon ausgeht, daß die Präferenzen der Egoistin durch $y - w < w - z$ gekennzeichnet sind, rechnet sie damit, daß der Verlust in der ersten Runde zu hoch ist und sich deshalb eine Signalstrategie für die Egoistin nicht lohnt. Die Konsequenz ist: Spielerin II interpretiert $a_I(t_1) = \text{"C"}$ als eindeutiges Signal für eine Altruistin, sie setzt $p_2 = 0$. Damit ergibt sich der Erwartungsnutzen für Spielerin II in t_2 durch:

$$(11) \quad \text{Eu}_{II}(t_2) = \begin{cases} a & \text{if } a_{II}(t_2) = \text{"C"} \\ b & \text{if } a_{II}(t_2) = \text{"D"} \end{cases}$$

⁷⁹Diese Annahme ist natürlich problematisch. Deshalb werden anschließend die Ergebnisse für einen allgemeineren Fall vorgestellt: mit Wahrscheinlichkeit α geht Spielerin II von der tatsächlichen Größenrelation aus, mit der Gegenwahrscheinlichkeit $1 - \alpha$ von der falschen.

Daraus folgt: "C" \succ "D".

$$\text{Fall 2: } a_I(t_1) = \text{"D"}$$

Da eine Altruistin niemals "D" spielen wird und es sich nach der Vermutung von Spielerin II für eine Egoistin nicht lohnt "C" zu spielen, geht Spielerin II mit Sicherheit davon aus, daß sie es mit einer Egoistin zu tun hat, sie setzt $p_2 = 1$ und ermittelt folgenden Erwartungsnutzen:

$$(12) \quad Eu_{II}(t_2) = \begin{cases} d & \text{if } a_{II}(t_2) = \text{"C"} \\ c & \text{if } a_{II}(t_2) = \text{"D"} \end{cases}$$

Daraus ergibt sich: "D" \succ "C".

Aufgrund der falschen Erwartungen von II gelingt es der Egoistin durch die Wahl von $a_{IE}(t_1) = \text{"C"}$ die Vermutungen ($p_2, 1-p_2$) der Spielerin II so zu beeinflussen, daß II in t_2 "C" spielt.

Das **sequentielle Gleichgewicht** in diesem Spiel lautet somit:

Das Vermutungssystem μ_{II} von Spielerin II ist gegeben durch:

$$(13) \quad \begin{aligned} & (p_1, 1-p_1) \text{ mit } \frac{a-b}{a-b+c-d} < p_1 < 1 \\ & (p_2, 1-p_2) = \begin{cases} (0,1) & \text{if } a_I(t_1) = \text{"C"} \\ (1,0) & \text{if } a_I(t_1) = \text{"D"} \end{cases} \end{aligned}$$

Das Strategieprofil $\{\Phi_1^*\}$ ist gegeben durch:

$$\Phi_{IE}^* = \{a_{IE}(t_1); a_{IE}(t_2)\} = \{C; D\}$$

$$\Phi_{II}^* = \{a_{II}(t_1); a_{II}(t_2)\} = \{D; C\}$$

$$\Phi_{IA}^* = \{a_{IA}(t_1); a_{IA}(t_2)\} = \{C; C\}$$

In diesem sehr vereinfachten Modellrahmen läßt sich also Kooperation bei einer Spielerin mit "D" als dominanter Strategie im stage-game bei endlicher (hier einmaliger) Wiederholung durch die Signalfunktion ihrer Aktion begründen. Es existiert ein se-

quentielles Gleichgewicht mit $a_{IE}(t_1) = "C"$, wobei die pay-off-Werte einer zusätzlichen Restriktion (9) genügen müssen. Durch die Wahl von "C" in der ersten Runde wird Spielerin II zu der Vermutung gebracht, daß die Präferenzen ihrer Mitspielerin "C" eindeutig bevorzugen und ihre Gegenspielerin somit altruistisch gesinnt ist. Ausschlaggebend für diese Tatsache war natürlich der irrtümliche Glaube von Spielerin II, daß sich die Signalstrategie für eine Egoistin nicht lohnen würde. Deshalb wird im nächsten Abschnitt diese harte Annahme fallengelassen und - wie bereits in Fußnote 8 erwähnt - eine allgemeinere Modellversion präsentiert.

Bevor wir diesen Weg beschreiten sei jedoch noch zu einer kurzen *methodologischen Exkursion* eingeladen. Die Verwendung spieltheoretischer Modelle öffnet den Blick auf wechselseitige Abhängigkeiten, Informationszustände und innere Überlegungen der beteiligten Spielerinnen. Die Methodik gestattet zudem durch die Wahl eines Gleichgewichtskonzepts eine jeweilige Vorstellung rationaler Gedankenführung herauszustellen. Diese an sich positiv zu bewertenden Eigenschaften werden durch eine Form der Kritik jedoch oftmals deutlich relativiert. Die Informationsverarbeitungsfähigkeiten würden unrealistisch hoch angesetzt, Menschen verhielten sich in vielen Entscheidungssituationen eher nach bewährten Daumenregeln als daß sie bereit oder in der Lage wären, zeitintensive und kognitiv anspruchsvolle Denkprozesse zu durchlaufen. Was bedeutet diese Kritik für unser Grundmodell, wenn bereits die einfachste Modellierung eines schlichten Aspekts menschlicher Urteilsbildung schwer nachvollziehbare Gedankengänge beinhaltet? Die Relevanz der Kritik hängt sicherlich davon ab, in welchem Kontext solche Überlegungen ablaufen. Es ist nicht die Methode, sondern eher eine ungeschickte Anwendung, die beim Modell der "kooperierenden Egoistin" zu Irritationen führen kann.

Um diesen Punkt zu verdeutlichen möchte ich das Grundmodell vor einem neuen Hintergrund präsentieren. Die Variation besteht hauptsächlich in der Motivation der Überlegungen von Spielerin II. Man kann beispielsweise die Informationen von II über die pay-offs von Spielerin I beschränken, was bedeuten würde, daß II nicht in der Lage ist, die oben vorgestellten Gedankenvorgänge durchzuführen. Wenn II nur weiß, daß sie entweder mit einer Egoistin oder einer Altruistin spielt, ergibt sich die Begründung von (7) wesentlich einfacher. Mit p_1 geht II davon aus auf eine Egoistin zu treffen, die für II - in Ermangelung weiterer Informationen - dadurch gekennzeichnet ist, daß "D" für sie dominante Strategie ist. Der Weg über (8) ist in diesem Modellrahmen nicht mehr notwendig. Spielerin II berücksichtigt in t_2 eine einfache Daumenregel: Erhält sie in t_1 das Signal $a_1(t_1) = "C"$, so geht sie davon aus mit einer Altruistin zu spielen und setzt entsprechend $p_2 = 0$; für $a_1(t_1) = "D"$ umgekehrt $p_2 = 1$. (11) und (12) behalten, allerdings anders motiviert, ihre Gültigkeit. Für die Egoistin, die in diesem Rahmen

einen zusätzlichen Informationsvorsprung besitzt, ist der Erfolg ihrer Signalstrategie garantiert. Ob sie $a_I(t_1) = "C"$ wählt, hängt nur noch davon ab, ob (9) erfüllt ist. Diese Variation sollte als Beispiel dafür verstanden werden, welche Konsequenzen unterschiedliche Informationsannahmen auf die durch ein bestimmtes Gleichgewichtskonzept erforderlichen kognitiven Abläufe besitzen.

Nach diesem Exkurs kehre ich zurück zum Grundmodell und diskutiere im nächsten Abschnitt die Auswirkungen einer allgemeineren Modellierung der Unsicherheit bei Spielerin II über die Größenordnung der pay-offs.

3. Variation 1: Unsicherheit über die pay-offs der Egoistin

Im Gegensatz zum bisher betrachteten Grundmodell gelte jetzt: Spielerin II geht mit einer subjektiven Wahrscheinlichkeit α davon aus, daß die Größenrelation $y - w > w - z$ zutrifft, was auch der Realität entspricht, mit der Gegenwahrscheinlichkeit $1 - \alpha$ vermutet sie: $y - w < w - z$ ⁸⁰. Im ersten Fall rechnet Spielerin II mit der Möglichkeit auf eine Egoistin zu treffen, für die es sich lohnt, durch die Wahl von $a_{IE}(t_1) = "C"$ eine Altruistin vorzutauschen, im zweiten Fall bleiben wir in der Vorstellungswelt des Grundmodells. Durch diese Modellerweiterung wird die Veränderung der a-priori-Wahrscheinlichkeiten durch bayesianisches Aufdatieren deutlich. Außerdem erhalten wir im Ergebnis eine zusätzliche Restriktion. Kooperation läßt sich nur dann begründen, wenn α nicht zu groß ist. Der "Spielraum" für eine gelungene Signalisierung wird immer enger, je stärker Spielerin II davon ausgeht, daß eine "Täuschungsstrategie" im Interesse der Egoistin liegt. Die Modellpräsentation verläuft analog zum Grundmodell: es wird aus der Sicht von Spielerin II argumentiert, in deren Überlegungen das Kalkül der Egoistin eingeht. Anschließend wird das Gleichgewicht für diesen Fall vorgestellt.

Wir beginnen mit den Überlegungen von Spielerin II in t_1 . In $(1 - p_1)$ -Fällen findet das "Altruistenspiel" statt. Wie zuvor - da eine Altruistin annahmegemäß "C" als dominante Strategie besitzt, erhält Spielerin II bei eigener Wahl von "D" als pay-off b , bei Wahl von "C" hingegen den höheren Wert a . In diesem Fall wäre es somit sinnvoll zu kooperieren.

Trifft Spielerin II jedoch auf eine Egoistin, wird die Überlegung wieder diffiziler. Zuerst muß II klären, welche Aktion die Egoistin wählen wird. Für diese ist entscheidend, welche Reaktion sie von II in der nächsten Runde zu erwarten hätte, falls sie $a_{IE}(t_1) = "C"$ wählen würde. Wird II dieses Signal dahingehend interpretieren, daß eine Altruistin vor ihr steht und deswegen in t_2 auf die von der Egoistin gewünschte Aktion "C" überwechseln? Dies ist davon abhängig, inwieweit II davon ausgeht, daß (9) erfüllt ist.

⁸⁰ Auch diese Annahme bleibt, wiewohl abgeschwächt im Vergleich zum Grundmodell, problematisch. Denn wieso sollte eine Spielerin falsche Erwartungen besitzen? Deswegen wird im nächsten Abschnitt eine weitere Variante präsentiert, bei der α nicht mehr als subjektive Wahrscheinlichkeit interpretiert wird, mit denen II glaubt, auf eine Egoistin zu treffen, für die sich eine Signalstrategie lohnt, sondern als tatsächlich vorhandener Anteil derartiger Egoistinnen.

Mit Wahrscheinlichkeit $1-\alpha$ vermutet sie das Gegenteil, "C" in t_1 würde von II daher als eindeutiges Signal für eine Altruistin aufgefaßt, was bedeutet, daß II daraufhin in Runde 2 "C" wählen würde. Für die Erwartungsbildung der II in t_1 ist an dieser Stelle relevant, daß sie mit $a_{IE}(t_1) = "D"$ rechnet, weil sich ja gemäß ihren Vermutungen eine Signalstrategie nicht lohnen würde.

Mit der Gegenwahrscheinlichkeit α jedoch geht sie davon aus, auf eine Egoistin zu treffen, für die sich das Signal "C" in t_1 auszahlen würde. Die Frage ist nun, welche Konsequenzen dieser Informationsstand auf das Verhalten von II in der nächsten Runde besitzt. Obwohl II in diesem Fall davon ausgeht, daß (9) gilt, ist sie nicht in der Lage, zu Beginn der nächsten Runde aus $a_{II}(t_1) = "C"$ eindeutig ihre Gegenspielerin erkennen zu können.

Dieser Umstand ist ursächlich dafür verantwortlich, warum sich Kooperation durch die Signalfunktion einer Aktion in einem endlich wiederholten Spiel bei einer Spielerin mit "D" als dominanter Strategie im stage-game begründen läßt.

Der nächste Schritt besteht also darin, sich die Überlegungen von II zu Beginn von t_2 zu verdeutlichen. In Abhängigkeit von $a_{IE}(t_1)$ ergibt sich $Eu_{II}(t_2)$. Deswegen wird im Rahmen einer Fallunterscheidung weiter argumentiert.

$$\text{Fall 1: } a_{IE}(t_1) = "C"$$

In einem ersten Schritt werden die Wahrscheinlichkeiten (p_2 , $1 - p_2$) durch bayesianisches Aufdatieren ermittelt, in einem zweiten die zugehörigen pay-offs. Liegt $a_I(t_1) = "C"$ vor, stellt sich die Situation für II wie folgt dar: die Mitspielerin kann entweder Altruistin (Wahrscheinlichkeit $1 - p_1$) oder aber Egoistin sein, für die sich ein Signal "C" lohnt, d.h. deren Präferenzen die Ungleichung $y - w > w - z$ erfüllen (Wahrscheinlichkeit αp_1). Die Wahrscheinlichkeit für das Auftreten von $a_I(t_1) = "C"$ beträgt somit $\alpha p_1 + 1 - p_1$. Die Wahrscheinlichkeit p_2 für eine Egoistin ergibt sich deshalb als

$$(14) \quad p_2 = \frac{\alpha p_1}{\alpha p_1 + 1 - p_1}$$

und die Gegenwahrscheinlichkeit $1 - p_2$ für eine Altruistin entsprechend als

$$(15) \quad 1 - p_2 = \frac{1 - p_1}{\alpha p_1 + 1 - p_1}$$

Da II damit rechnet, daß eine Egoistin in Runde 2 auf jeden Fall "D" spielen wird und eine Altruistin ihre dominante Strategie "C" wählt, erhält man

$$(16) \quad Eu_{II}(t_2) = \begin{cases} \frac{\alpha p_1}{\alpha p_1 + 1 - p_1} d + \frac{1 - p_1}{\alpha p_1 + 1 - p_1} a & \text{if } a_{II}(t_2) = "C" \\ \frac{\alpha p_1}{\alpha p_1 + 1 - p_1} c + \frac{1 - p_1}{\alpha p_1 + 1 - p_1} b & \text{if } a_{II}(t_2) = "D" \end{cases}$$

Ein Vergleich beider Erwartungswerte ergibt, daß "C" \succ "D" dann und nur dann gilt, wenn

$$(17) \quad \alpha < \frac{(1 - p_1)(a - b)}{p_1(c - d)}$$

erfüllt ist.

Wenn (17) gilt und beiden Spielerinnen bekannt ist, wird II damit rechnen, daß eine Egoistin, von der sie vermutet, daß (9) erfüllt ist, in der ersten Runde "C" spielt und damit zur "täuschenden" Egoistin wird.

Zur Interpretation von (17): Je geringer α ist, d.h. je weniger II damit rechnet, daß es für I_E lohnend ist, in t_1 "C" zu spielen, desto stärker signalisiert $a_{II} = "C"$ eine Altruistin. Anders betrachtet: Je größer die a-priori-Vermutung p_1 und je größer der "Verlust" $w - z$ in der ersten Runde ist, umso geringer muß ceteris paribus die Vermutung α sein. Für die Egoistin ist dadurch der Erfolg ihrer Signalstrategie (nämlich in t_1 "C" zu spielen und Spielerin II dadurch zu veranlassen in t_2 ebenfalls "C" zu spielen) im Unterschied zum Grundmodell nicht mehr sicher.

$$\text{Fall 2 : } a_I(t_I) = "D"$$

"D" ist ein eindeutiges Signal für eine Egoistin, deswegen gilt: $p_2 = 1$. Der Erwartungsnutzen in diesem Fall lautet wieder (12). Daraus folgt: "D" \succ "C".

An dieser Stelle schließt sich der Kreis. Spielerin II kann nun ihren Erwartungsnutzen in t_1 formulieren. Im Unterschied zu (7) gilt nun:

$$(18) \quad \text{Eu}_{II}(t_1)^{81} = \begin{cases} p_1 \alpha a + p_1 (1 - \alpha) d + (1 - p_1) a & \text{if } a_{II}(t_1) = "C" \\ p_1 \alpha b + p_1 (1 - \alpha) c + (1 - p_1) b & \text{if } a_{II}(t_1) = "D" \end{cases}$$

Noch einmal die Begründung für (18): die Altruistin (wird von II in $(1-p_1)$ -Fällen erwartet) wählt in t_1 "C"; eine Egoistin, von der II glaubt, daß (9) für sie nicht erfüllt ist (wird in $p_1(1-\alpha)$ -Fällen erwartet), wählt in t_1 "D" und schließlich: eine Egoistin, von der II glaubt, daß (9) für sie erfüllt ist (wird in den restlichen $p_1\alpha$ -Fällen erwartet), wählt in t_1 "C". Die pay-offs können wiederum anhand der Normalform aus Abb.1 nachvollzogen werden. Die erste Aktion von II hängt davon ab, wie hoch die a-priori-Wahrscheinlichkeit p_1 in Relation zu ihren pay-offs ist. Für $a_{II}(t_1) = "D"$ muß (im Unterschied zu (10)) gelten:

$$(19) \quad p_1 > \frac{a - b}{(1 - \alpha)(a - b + c - d)}$$

Diese Ungleichung läßt sich einfach interpretieren: bei gegebenen pay-offs und gegebener Vermutung α wird II in t_1 umso eher "D" wählen, je weniger sie damit rechnet, mit einer Altruistin zu spielen, d.h. je größer p_1 ist. Mit einem anderen Blickwinkel: bei gegebenen pay-offs und gegebenen a-priori-Wahrscheinlichkeiten wird II umso eher "D" wählen, je weniger sie vermutet auf eine "täuschende" Egoistin zu treffen, d.h. je kleiner α ist.

Für die Egoistin gilt in t_1 wieder (8), da sie $a_{II}(t_1) = "D"$ aufgrund von (19) berücksichtigt. Im Gegensatz zum Grundmodell kann sich die Egoistin jetzt jedoch nicht mehr sicher sein, daß sie Spielerin II durch Wahl von $a_I(t_1) = "C"$ dazu veranlassen wird, ihrerseits in t_2 "C" zu spielen. Wie vorher diskutiert, hängt die optimale Reaktion von II davon ab, welches α sie besitzt. Deshalb ist bei der Ermittlung von $u_{IE}(t_1) + u_{IE}(t_2)$ eine neuerliche Fallunterscheidung notwendig. Im weiteren wird berücksichtigt, daß I_E α kennt.

⁸¹(7) erhält man aus (18), wenn man $\alpha = 0$ setzt.

$$\text{Fall 1: } \alpha < \frac{(1-p_1)(a-b)}{p_1(c-d)}$$

In diesem Fall gilt $a_{II}(t_2) = "C"$ if $a_I(t_1) = "C"$. Daraus ergeben sich folgende Nutzenwerte⁸²:

Φ_{IE}	$\{ D , D \}$	$\{ C , D \}$
u_{IE}	$w + w$	$z + y$

Die Situation ist identisch mit der im Grundmodell skizzierten. Deshalb gilt: $\Phi_{IE}^* = \{C, D\}$. Für die angegebenen Parameterkonstellationen bzgl. der Wahrscheinlichkeiten p_1 und α läßt sich somit Kooperation der Egoistin in der ersten Runde begründen.

$$\text{Fall 2: } \alpha > \frac{(1-p_1)(a-b)}{p_1(c-d)}$$

In diesem Fall gilt: $a_{II}(t_2) = "D"$ if $a_I(t_1) = "C"$. Daraus ergeben sich folgende Nutzenwerte:

Φ_{IE}	$\{ D , D \}$	$\{ C , D \}$
u_{IE}	$w + w$	$z + w$

⁸²Da die Aktion der Egoistin in t_2 keine Signalfunktion mehr besitzt, kann man den Nutzenvergleich auf die angegebenen Strategien beschränken.

Das Signal "C" verfehlt seinen Zweck, da Spielerin II in t_2 "D" wählt. Das Ergebnis ändert sich deutlich: die optimale Strategie Φ_{IE}^* lautet jetzt $\{ D, D \}$. Bei hinreichend großem α kann man Kooperation also nicht mehr begründen.

Das sequentielle Gleichgewicht in dieser Modellvariante lautet:

Das Vermutungssystem μ_{II} von Spielerin II ist gegeben durch:

$$(p_1, 1 - p_1) \text{ mit } \frac{a-b}{(1-\alpha)(a-b+c-d)} < p_1 < 1$$

$$(p_2, 1 - p_2) = \begin{cases} \left(\frac{\alpha p_1}{\alpha p_1 + 1 - p_1}, \frac{1 - p_1}{\alpha p_1 + 1 - p_1} \right) & \text{if } a_I(t_1) = "C" \\ \left(1, 0 \right) & \text{if } a_I(t_1) = "D" \end{cases}$$

(20)

Das Strategieprofil $\{ \Phi_i^* \}$ ist gegeben durch:

$$\Phi_{IE}^* = \{ a_{IE}(t_1); a_{IE}(t_2) \} = \begin{cases} \{C, D\} & \alpha < \frac{(1-p_1)(a-b)}{p_1(c-d)} \\ \{D, D\} & \alpha > \frac{(1-p_1)(a-b)}{p_1(c-d)} \end{cases} \text{ if}$$

$$\Phi_{II}^* = \{ a_{II}(t_1); a_{II}(t_2) \} = \begin{cases} \{D, C\} & \alpha < \frac{(1-p_1)(a-b)}{p_1(c-d)} \\ \{D, D\} & \alpha > \frac{(1-p_1)(a-b)}{p_1(c-d)} \end{cases}$$

$$\Phi_{IA}^* = \{ a_{IA}(t_1); a_{IA}(t_2) \} = \{ C; C \}$$

Wieder läßt sich, wie bereits im Grundmodell, Kooperation begründen. Allerdings ist wegen der Möglichkeit, daß Spielerin II im Glauben sein kann, auf eine Egoistin zu treffen ($\alpha > 0$), für die sich eine Signalisierung von "C" lohnt, die zusätzliche Restriktion (17) nötig.

4. Variation 2: Zwei Egoistentypen⁸³

Im Unterschied zur ersten Variante des Grundmodells werden jetzt $(\alpha, 1 - \alpha)$ nicht mehr als subjektive Wahrscheinlichkeiten von Spielerin II interpretiert, sondern es wird davon ausgegangen, daß zwei unterschiedliche Egoistentypen *existieren*. α soll dabei den Anteil der Egoistinnen kennzeichnen, für die es sich lohnt durch $a_{IE}(t_1) = "C"$ eine Altruistin vorzutauschen (Typ E_s). Deren Präferenzen seien deshalb durch die Ungleichung

$$(21) \quad y_s - w_s > w_s - z_s$$

gekennzeichnet⁸⁴. Entsprechend sei der restliche Anteil der Egoistinnen $(1 - \alpha)$ durch

$$(22) \quad y - w < w - z$$

charakterisiert (Typ E_{ns}). Spielerin II rechnet also damit, in $(1-p_1)$ -Fällen im "Altruistenspiel" zu landen, in αp_1 -Fällen ins "signalisierende Egoistenspiel" zu geraten sowie sich in $(1-\alpha)p_1$ -Fällen im "nicht-signalisierenden Egoistenspiel" wiederzufinden. Die drei Spiele mit entsprechender Wahrscheinlichkeit lassen sich leicht anhand einer Abbildung veranschaulichen:

⁸³ Bis auf die angegebenen Modifikationen gelten weiterhin die Annahmen aus dem vorherigen Abschnitt.

⁸⁴ Die abweichende Notation ist notwendig, weil die beiden Egoistentypen unterschiedliche Präferenzen besitzen. Die Indexierung soll die signalisierende Egoistin verdeutlichen. Außerdem gelte $y_s > x_s > w_s > z_s$.

signalisierendes Egoistenspiel	αp_1	<table> <tr> <td>$\begin{array}{c c} \text{II} & \\ \hline \text{I}_E & \end{array}$</td><td>C</td><td>D</td></tr> <tr> <td>C</td><td>x_s, a</td><td>z_s, b</td></tr> <tr> <td>D</td><td>y_s, d</td><td>w_s, c</td></tr> </table>	$\begin{array}{c c} \text{II} & \\ \hline \text{I}_E & \end{array}$	C	D	C	x_s, a	z_s, b	D	y_s, d	w_s, c
$\begin{array}{c c} \text{II} & \\ \hline \text{I}_E & \end{array}$	C	D									
C	x_s, a	z_s, b									
D	y_s, d	w_s, c									
nicht-signalisierendes Egoistenspiel	$(1-\alpha)p_1$	<table> <tr> <td>$\begin{array}{c c} \text{II} & \\ \hline \text{I}_E & \end{array}$</td><td>C</td><td>D</td></tr> <tr> <td>C</td><td>x, a</td><td>z, b</td></tr> <tr> <td>D</td><td>y, d</td><td>w, c</td></tr> </table>	$\begin{array}{c c} \text{II} & \\ \hline \text{I}_E & \end{array}$	C	D	C	x, a	z, b	D	y, d	w, c
$\begin{array}{c c} \text{II} & \\ \hline \text{I}_E & \end{array}$	C	D									
C	x, a	z, b									
D	y, d	w, c									
Altruistenspiel	$1 - p_1$	<table> <tr> <td>$\begin{array}{c c} \text{II} & \\ \hline \text{I}_A & \end{array}$</td><td>C</td><td>D</td></tr> <tr> <td>C</td><td>$x+\beta a, a$</td><td>$z+\beta b, b$</td></tr> <tr> <td>D</td><td>$y+\beta d, d$</td><td>$w+\beta c, c$</td></tr> </table>	$\begin{array}{c c} \text{II} & \\ \hline \text{I}_A & \end{array}$	C	D	C	$x+\beta a, a$	$z+\beta b, b$	D	$y+\beta d, d$	$w+\beta c, c$
$\begin{array}{c c} \text{II} & \\ \hline \text{I}_A & \end{array}$	C	D									
C	$x+\beta a, a$	$z+\beta b, b$									
D	$y+\beta d, d$	$w+\beta c, c$									

Abb.2: Spiele bei zwei Egoistentypen

Wie sieht nun die Entscheidungssituation für Spielerin II aus?

Zu Beginn von t_1 geht II mit $(1 - p_1)$ davon aus, auf eine Altruistin zu treffen, die durch "C" als dominante Strategie gekennzeichnet ist. Deswegen ergibt sich für II ein pay-off von b bei eigener Wahl von "D" bzw. a bei eigener Wahl von "C".

Die nächste Möglichkeit für II besteht darin, sich mit einer "nicht signalisierenden" Egoistin auseinanderzusetzen zu müssen, was in $(1-\alpha)p_1$ -Fällen von II erwartet wird. Um ihren eigenen pay-off bestimmen zu können, muß sich II in die Lage von $I_{E_{NS}}$ versetzen. Für diese lohnt sich aufgrund von (22) die Wahl von $a_{I_{E_{NS}}}(t_1) = "C"$ nicht, da selbst bei erfolgreicher Signalisierung der Nutzenverlust in t_1 größer als der Nutzen Gewinn in t_2 wäre. Deshalb wird eine derartige Egoistin in t_1 immer "D" wählen, was dazu führt, daß II mit einem pay-off von c bei eigener Wahl von "D" bzw. d bei eigener Wahl von "C" rechnet.

Mit der restlichen Wahrscheinlichkeit αp_1 wird II an eine "signalisierende Egoistin" I_{ES} geraten. Bei deren Wahl zwischen "C" und "D" in t_1 kommt es entscheidend darauf an, in welchem Maße sie mit einem Erfolg ihrer Signalstrategie rechnen kann. Aufgrund von (21) lohnt sich diese für I_{ES} . Obwohl Spielerin II (21) kennt, ist sie nicht in der Lage aus der Beobachtung von $a_{I_{ES}}(t_1) = "C"$ eindeutig zwischen einer signalisierenden Egoistin und einer Altruistin zu unterscheiden. Bei der Wahl zwischen "C" und "D" in t_2 sind für II nun $(p_2, 1-p_2)$ relevant (siehe (14) und (15) aus der ersten Variation) sowie die Erwartungsnutzenbildung (16). Eine Begründung von $a_{II}(t_2) = "C"$ ist wieder nur über (17) möglich. Die Interpretation von (17) verläuft allerdings in anderen Bahnen. Je geringer α , also der Anteil der Egoistinnen, für die sich eine Signalstrategie lohnt, desto eher wird es gerade diesen Egoistentypen möglich sein, erfolgreich eine Altruistin vorzutäuschen und damit Spielerin II zur Kooperation in der zweiten Runde zu bewegen. Wenn (17) erfüllt und beiden Spielerinnen bekannt ist, geht II davon aus, daß I_{ES} in der ersten Runde "C" wählt. Als Erwartungsnutzen für Spielerin II in t_1 ergibt sich wieder (18), für "D" \succ "C" muß (19) gelten. Daraus erhält man das Gleichgewicht (20) aus dem vorherigen Abschnitt.

Neben der im Abschnitt 3 diskutierten Verallgemeinerung der Vermutungen der uninformierten Spielerin über die Größenordnungen der pay-offs von Spielerin I und der alternativen Interpretation von α in diesem Abschnitt stellt die Beschränkung auf zwei Runden im Grundmodell eine weitere, erhebliche Vereinfachung dar. Deshalb wird im nächsten Abschnitt der Frage nachgegangen, welche Ergebnisse in einem n -Perioden-Spiel zu erwarten sind.

5. Variation 3: Ein n-Perioden-Modell

Dieser Abschnitt wird in einer etwas anderen Form präsentiert. Zuerst wird im Rahmen einer Proposition das sequentielle Gleichgewicht postuliert (1.) und anschließend die dafür notwendigen Bedingungen, sequentielle Rationalität der beiden Spielerinnen (2. und 3.) sowie Konsistenz (4.) erörtert.

Die Anforderung der sequentiellen Rationalität impliziert, daß keine der Spielerinnen, ausgehend von einer beliebigen Informationsmenge innerhalb des Spiels und gegeben das Strategieprofil $\{ \Phi_i^* \}$ und das Vermutungssystem μ_{II} , einen Anreiz hat von ihrer Gleichgewichtsstrategie abzuweichen. Konsistenz bedeutet schlicht und einfach, daß die Wahrscheinlichkeiten in späteren Perioden endogen aus den Aktionen und den Wahrscheinlichkeiten früherer Perioden abgeleitet werden müssen.

Das Urteil darüber, welche der alternativ möglichen Präsentationsweisen einsichtiger ist, bleibe der Leserin überlassen. Eine Annahme noch vorneweg: Anstelle der Vermutung α aus den Abschnitten 3 und 4 sei Spielerin II jetzt dadurch charakterisiert, daß sie mit einer Wahrscheinlichkeit π damit rechnet, daß

$$(23) \quad n > \frac{2x - y - z}{x - w} > 0$$

gilt⁸⁵ und mit der Gegenwahrscheinlichkeit $1 - \pi$ von der umgekehrten Größenrelation ausgeht⁸⁶. Der Unterschied in der Modellierung zwischen α und π ergibt sich dadurch, daß sich die Unsicherheit immer auf den vermuteten Signal-trade-off der Egoistin bezieht. Es ist von entscheidender Bedeutung, ob II glaubt, daß es sich für eine Egoistin lohnt "falsch" zu spielen oder nicht. Dieser trade-off wiederum ist in einem n-Perioden-Spiel anders als in einer zweiperiodischen Version. Die Bedeutung von Gleichung (23) wird bei der Diskussion der sequentiellen Rationalität der Egoistin deutlich. Damit aber zur Präsentation des Gleichgewichts.

⁸⁵Bei gegebenen pay-offs wird mit immer größerer Rundenzahl ($n \rightarrow \infty$) π gegen 1 gehen.

⁸⁶Auch bezüglich π sind die beiden Interpretationen aus den vorausgegangenen Abschnitten (hinsichtlich α) entsprechend anwendbar.

1.

Proposition: $\{\Phi_1^*\}$ und μ_{II} bilden ein sequentielles Gleichgewicht.

Das Vermutungssystem μ_{II} von Spielerin II ist gegeben durch:

$$\begin{aligned}
 (p_1, 1 - p_1) \quad & \text{mit } 0 < \frac{a-b}{(1-\pi)(a-b+c-d)} < p_1 < 1 \\
 (p_j, 1 - p_j) = & \begin{cases} \left(\frac{\pi p_1}{\pi p_1 + 1 - p_1}, \frac{1 - p_1}{\pi p_1 + 1 - p_1} \right) & \text{if } a_i(t_1) = "C" \\ (1, 0) & \text{if } a_i(t_1) = "D" \end{cases} \\
 & \text{für } j = 2, \dots, n \quad \text{mit } 0 < p_n < \frac{a-b}{(1-\pi)(a-b+c-d)} < 1
 \end{aligned}$$

(24)

Das Strategieprofil $\{\Phi_1^*\}$ ist gegeben durch:

$$\begin{aligned}
 \Phi_{IE}^* &= \{ a_{IE}(t_1); a_{IE}(t_2); \dots; a_{IE}(t_{n-1}); a_{IE}(t_n) \} \\
 &= \begin{cases} \{ C; C; \dots; C; D \} & n > \frac{2x-y-z}{x-w} \\ \{ D; D; \dots; D; D \} & \text{if } n < \frac{2x-y-z}{x-w} \end{cases} \\
 \Phi_{II}^* &= \{ a_{II}(t_1); a_{II}(t_2); \dots; a_{II}(t_{n-1}); a_{II}(t_n) \} \\
 &= \begin{cases} \{ D; C; \dots; C; C \} & n > \frac{2x-y-z}{x-w} \\ \{ D; D; \dots; D; D \} & \text{if } n < \frac{2x-y-z}{x-w} \end{cases} \\
 \Phi_{IA}^* &= \{ a_{IA}(t_1); a_{IA}(t_2); \dots; a_{IA}(t_{n-1}); a_{IA}(t_n) \} \\
 &= \{ C; C; \dots; C; C \}
 \end{aligned}$$

2. Sequentielle Rationalität von Spielerin II

Hier gilt die analoge Überlegung aus dem Grundmodell. Spielerin II überlegt sich ihre Aktion in einer beliebigen Periode prinzipiell immer in Abhängigkeit von der Aktion der Spielerin I in der Vorperiode. In t_1 wurde wieder die gleiche a-priori-Vermutung ($p_1, 1-p_1$) gesetzt. Es gilt daher: $a_{II}(t_1) = "D"$. In den Folgeperioden sind die erwarteten pay-offs sowohl vom Signal der Gegenspielerin (deren Aktion in der Vorperiode) als auch von den entsprechenden Wahrscheinlichkeiten abhängig. Für $j = 2, \dots, n-1$ ist somit eine Fallunterscheidung vorzunehmen:

Fall 1: $a_I(t_{j-1}) = "C"$

in t_2 gilt:

$$(25) \quad Eu_{II}(t_2) = \begin{cases} \frac{\pi p_1}{\pi p_1 + 1 - p_1} b + \frac{1 - p_1}{\pi p_1 + 1 - p_1} b = b & a_{II}(t_2) = "D" \\ \frac{\pi p_1}{\pi p_1 + 1 - p_1} a + \frac{1 - p_1}{\pi p_1 + 1 - p_1} a = a & \text{if } a_{II}(t_2) = "C" \end{cases}$$

Wählt Spielerin I in t_1 "C", so wird II bei der Ermittlung ihres erwarteten pay-offs in t_2 die angegebenen Wahrscheinlichkeiten verwenden⁸⁷. Bei eigener Wahl von "C" ergibt sich: Die Wahrscheinlichkeit für das Auftreten von $a_I(t_1) = "C"$ beträgt $\pi p_1 + 1 - p_1$. Mit $p_2 = \pi p_1 / (\pi p_1 + 1 - p_1)$ rechnet II damit auf eine Egoistin zu treffen, die selbst "C" wählt. Als pay-off resultiert a. In $(1-p_2)$ -Fällen trifft II auf eine Altruistin, die Φ_{IA}^* spielt und damit ebenfalls "C". Der pay-off in diesem Fall beträgt wieder a. Eu_{II} bei Wahl von "D" ergibt sich analog.

Durch die Wahl von $a_I(t_1) = "C"$ ergibt sich für Spielerin II ein zusätzlicher Informationsgewinn: Wußte sie zu Beginn des Spiels nicht, ob sie einer Altruistin oder einer Egoistin, für die sich eine Signalstrategie lohnt oder aber einer Egoistin, für die sich eine solche Signalstrategie nicht lohnt, gegenüber steht, so kann sie zu Beginn von t_2 den letzten Fall mit Sicherheit ausschließen. Dieser zusätzliche Erkenntnisgewinn schlägt sich in der Revidierung ihrer ursprünglichen Vermutung durch bayesianisches Aufdatieren nieder. Da allgemein $a > b$ gilt, folgt: $a_{II}(t_2) = "C"$. In den folgenden Perioden t_i für $i = 3, \dots, n-1$ gilt:

Spielerin II erhält durch die Aktion von Spielerin I in der vorhergehenden Periode jetzt keine zusätzliche Information. Sowohl eine Altruistin als auch eine "täuschende" Egoistin (siehe dazu die anschließende Betrachtung aus der Sicht der Egoistin) wählen

⁸⁷Zur Ermittlung dieser Wahrscheinlichkeiten vgl. auch analog (14) und (15).

eine dauerhafte Kooperationsstrategie in den ersten $n-1$ Perioden. Spielerin II wird deswegen $(p_2, 1-p_2)$ verwenden. Es gilt daher: $Eu_{II}(t_1) = Eu_{II}(t_2)$.

In der letzten Periode bilden wieder $(p_2, 1-p_2)$ die Grundlage, es ändern sich jedoch die erwarteten pay-offs, da Spielerin II hier mit Sicherheit davon ausgeht, daß eine Egoistin - unabhängig von n - immer "D" wählen wird. Ein Anreiz durch "C" Altruismus zu signalisieren besteht in t_n nicht mehr. Als Erwartungsnutzen resultiert:

$$(26) \quad Eu_{II}(t_n) = \begin{cases} p_2 d + (1-p_2)a & \text{if } a_{II}(t_n) = "C" \\ p_2 c + (1-p_2)b & \text{if } a_{II}(t_n) = "D" \end{cases} \text{ mit } p_2 = \frac{\pi p_1}{\pi p_1 + 1 - p_1}$$

Ein Vergleich der beiden Erwartungswerte zeigt, daß $a_{II}(t_n) = "C"$ gilt, falls

$$(27) \quad p_2 < \frac{a-b}{a-b+c-d} \quad (\text{"final-round-condition"}).$$

gilt.

Anmerkung zur Interpretation der "final-round-condition": Je geringer der Unterschied zwischen c und d ist, desto mehr nähert sich der Bruch dem Wert 1 und desto allgemeiner wird diese Bedingung erfüllt. Eine Annäherung von d an c bedeutet aus Sicht von Spielerin II aber auch einen entsprechend geringeren Verlust durch die Wahl von "C", wenn *irrtümlicherweise* "C" als Aktion der Gegenspielerin unterstellt wurde. Die Wahl von "D" als erste Aktion von II in t_1 wird bei einer solchen Konstellation natürlich ebenfalls immer unwahrscheinlicher.

$$\text{Fall 2: } a_I(t_{j-1}) = "D"$$

Als Erwartungsnutzen für Spielerin II in t_j erhält man:

$$(28) \quad Eu_{II}(t_j) = \begin{cases} c & \text{if } a_{II}(t_j) = "D" \\ d & \text{if } a_{II}(t_j) = "C" \end{cases}$$

Begründung: Bei Erhalt des Signals "D" setzt Spielerin II die Egoistenwahrscheinlichkeit $p_j = 1$, da eine Altruistin niemals "D" spielen würde. Dadurch ergeben sich die pay-offs aus der Normalform des Egoistenspiels wie angegeben. Spielerin II wählt "D", da $c > d$ gilt.

3. Sequentielle Rationalität der Egoistin

Zu Beginn des Spiels (in t_1) läßt sich die Argumentation aus dem Grundmodell fortführen. Die Wahl von "C" bei gegebener Wahl von $a_{II}(t_1) = "D"$ bedeutet einerseits einen Verzicht von $w - z$ in t_1 , andererseits wird dadurch Spielerin II veranlaßt in den anschließenden Runden ebenfalls "C" zu wählen. Darauf mit "D" zu antworten, würde zwar den pay-off für die Egoistin in der entsprechenden Runde von x auf y erhöhen, gleichzeitig würde Spielerin II allerdings als Reaktion auf dieses Signal die Egoistin "entlarven" und von $p_j = 1$ ausgehen. Für den Rest der Spieldauer würde sich ständige beidseitige Nichtkooperation ergeben. Die unterschiedlichen Erträge verschiedener Strategien werden anhand eines Vergleichs von drei alternativ möglichen Strategien verdeutlicht. Bei der Ermittlung des pay-offs in der letzten Runde wird davon ausgegangen, daß die "final-round-condition" (27) gilt und deshalb $a_{II}(t_n) = "C"$.

	[1]	[2]	[3]
Φ_{IE}	{C,C,...,C,C,D}	{D,D,...,D,D,D}	{C,C,...,D,D,D}
u_{IE}	$z+x+...+x+x+y =$ $z+(n-2)x+y$	$w+w+.....+w =$ nw	$z+x+...+y+w+w =$ $z+(i-2)x+y+(n-i)w$

Abb.3: pay-offs der Egoistin im n-Perioden-Spiel

Die dritte Strategie sieht Beginn mit Kooperation vor, in Runde i wird auf "D" gewechselt und bis zum Ende in t_n beibehalten. Ein Vergleich zwischen [1] und [3] zeigt, daß $u_{IE}(\Phi_{IE}=[1]) > u_{IE}(\Phi_{IE} = [3])$ für $n > i$ gilt. Eine teilweise Kooperation ist einer (ausschließlich der letzten Runde) dauerhaften Kooperation unterlegen. Damit [1] auch gegenüber [2] vorzuziehen ist, muß (23) gelten. Eine Egoistin wird also nur dann, wenn ihre pay-offs diese Ungleichung bei gegebener Periodenzahl n erfüllen, die Strategie [1] wählen⁸⁸. An dieser Stelle sollte die Notwendigkeit der eingangs gemach-

⁸⁸ Variiert man die "final-round-condition", d.h. geht man davon aus, daß $p_2 < (a-b)/(a-b+c-d)$ gilt, resultiert folgendes Ergebnis: Spielerin II wird in der letzten Runde nicht kooperieren, dadurch sinkt der pay-off der

ten Annahme über π deutlich werden. Da Spielerin II das Kalkül der Egoistin nachvollziehen kann, ist der Erfolg der Signalisierung wie auch unter Abschnitt 3 und 4 davon abhängig, inwieweit II glaubt, daß es für die Egoistin rational ist "falsch" zu spielen.

4. Konsistenz

Die a-priori-Vermutung $(p_1, 1-p_1)$ wird wieder exogen gesetzt. Die Wahrscheinlichkeiten $(p_2, 1-p_2)$ in der folgenden Runde ergeben sich hingegen endogen in Abhängigkeit vom erhaltenen Signal - der Aktion der Spielerin I in der ersten Periode. Für den Fall $a_1(t_1) = "D"$ wird II mit Sicherheit davon ausgehen, daß Spielerin I Egoistin ist, da "C" dominante Strategie im stage-game für eine Altruistin ist und für diese keine Signalfunktion ihrer Aktion besteht. Bei $a_1(t_1) = "C"$ kann Spielerin II auf der Grundlage ihrer zur Verfügung stehenden Information dagegen nicht eindeutig erkennen, welchen Typ ihre Gegenspielerin besitzt. I kann entweder eine Altruistin sein, die ihre dominante Strategie spielt oder aber eine Egoistin, für die sich das Signal "C" lohnt. Eine Altruistin wird in der zweiten Runde mit der Häufigkeit $1-p_1$ erwartet, eine "signalisierende Egoistin" mit πp_1 . Die Wahrscheinlichkeit, daß der entsprechende Knoten in der Informationsmenge der Spielerin II in t_2 überhaupt erreicht wird, beträgt somit $1-p_1+\pi p_1$. Durch bayesianisches Aufdatieren erhält man schließlich die in der Proposition angegebenen Wahrscheinlichkeiten $(p_2, 1-p_2)$. In den folgenden Perioden werden diese Wahrscheinlichkeiten weiterhin bei der Erwartungsbildung zugrundegelegt, da wie bereits erwähnt die Basis für eine Anpassung - zusätzliche neue Information - fehlt.

Egoistin in t_n bei Wahl von Strategie [1] von y auf w. Die signalbedingte Kooperation ist jetzt begründbar, falls $n > (2x-w-z)/(x-w)$ gilt. Die Restriktion verschärft sich, was auch intuitiv zu erwarten war.

VERSUCH ÜBER DIE IDEE DER BINDUNG

(Zum Risiko einer commitment-Strategie in einem Problem der Koordination)

1. Bindung und Risiko

Traditionell wird zur Lösung von Chicken oder Battle-of-Sexes Strukturen des öfteren die Möglichkeit der freiwilligen Selbstbindung eines Spielers auf einer dem eigentlichen Spiel vorgelagerten Ebene diskutiert. Die Form der Selbstbindung kann entweder in der Wahl einer bestimmten Technologie, des zur Verfügung stehenden Einkommens oder der Präferenzen zum Ausdruck kommen.

Das zentrale Anwendungsbeispiel findet sich im Problem der Zeitinkonsistenz. Eine Regierung verzichtet auf die Entscheidungsbefugnis über die Geldpolitik zugunsten einer unabhängigen Notenbank um Glaubwürdigkeit und dadurch Wirksamkeit zu erreichen⁸⁹. Wie auch immer die Selbstbindung formuliert ist, zentral für die Lösung eines solchen Dilemmas sind Unterschiede der beiden Spieler hinsichtlich ihrer Bindungsmöglichkeiten und -fähigkeiten. Gelingt es durch eine wie auch immer geartete Aktion dem Gegenspieler glaubwürdig zu signalisieren, daß man selbst auf keinen Fall kooperieren wird, so resultiert als eindeutiges Nash-Gleichgewicht in einem Chicken-Spiel (D,C). Somit wird eine Bindungsstrategie vorteilhaft, wenn der Nutzen aus (D,C) abzüglich der Bindungskosten immer noch größer ist als der Nutzen, den der betreffende Spieler im eigentlichen Chickenspiel erreichen könnte.

Doch selbst wenn die Kosten einer Bindung bekannt sind, kann es von Nachteil sein sich vorzeitig eines Teils seiner Handlungsmöglichkeiten zu berauben. Die Schwierigkeiten können darin liegen, daß zu einem Zeitpunkt, zu dem sich der Spieler entscheiden muß, ob er eine prinzipiell mögliche Bindung eingeht, noch nicht klar ist, welche Interessenskonstellation in der nächsten Periode vorliegt. Anders ausgedrückt: es be-

⁸⁹ Vgl. dazu z.B. Persson/Tabellini (1990) und Kapitel 4 in Hargreaves Heap (1992a).

steht Unsicherheit über die Veränderung der pay-offs im Vergleich zum heutigen Zeitpunkt.

Diese Unsicherheit motiviere ich in den Modellen dieses Kapitels unter der Berücksichtigung eines globalen Umweltproblems⁹⁰ (zum einen) sowie eines psychologischen Faktors (zum anderen).

Internationale umweltpolitische Entscheidungen stehen im Grunde sehr häufig vor der Schwierigkeit, langfristige Konsequenzen zu verursachen. Dieses Problem wird oftmals dadurch verschärft, daß konkurrierende naturwissenschaftliche Theorien zu gravierend unterschiedlichen Prognosen über das Ausmaß anthropogener Eingriffe in das Ökosystem "Erde" führen.

In einem psychologischen Kontext interpretiert, läßt sich der Grundgedanke folgendermaßen fassen: ein Akteur mit der Option, sich in Runde 1 an eine Handlung zu binden, weiß, daß seine Bindungsoption die Fairnessgefühle seines Gegenspielers beeinflusst. Er ist sich aber nicht im klaren darüber, mit welcher Intensität dies geschieht.

⁹⁰ vgl. dazu die Ausführungen bei Althammer/Buchholz (1993), v.a. S.298.

2. Ein Beispiel aus der Umweltökonomie

In diesem Abschnitt wird die Problematik der weltweiten Reduktion von CO_2 (1.) als empirischer Hintergrund für die Modellierung eines einfachen zweistufigen Modells (2.) verwendet.

1.

Eine der Hauptursachen für das Scheitern verbindlicher globaler Vereinbarungen in der Treibhausproblematik liegt in den Unsicherheiten bzgl. der zu erwartenden Kosten eines CO_2 -Anstiegs. Das Intergovernmental Panel on Climatic Change (IPCC)⁹¹, 1990 im Auftrag der World Meteorological Organization und des UN Environmental Programs veröffentlicht, stellte hierzu neuere Erkenntnisse vor.

Eine Verdoppelung des CO_2 -Bestandes führt zu einer Erhöhung der jährlichen Durchschnittstemperatur zwischen $1,5^\circ\text{C}$ und $4,5^\circ\text{C}$. Auf der Grundlage von General Circulation Models (GCM) wurde versucht, die Auswirkungen von Temperaturänderungen auf das Klima zu prognostizieren. Allein die Unsicherheit über die feedback-Wirkungen von Wolken lieferte signifikante Unterschiede in den Ergebnissen der einzelnen GCM's hinsichtlich der zu erwartenden Änderungen verschiedener Klimavariablen wie Wind, Luftfeuchtigkeit oder Niederschläge. Sämtliche der möglichen Konsequenzen sind zudem sowohl räumlich als auch zeitlich unsicher. Ein Land, das im Glauben einer Chickenstruktur gegenüberzustehen in internationalen Verhandlungen den "hardliner" spielt, um den vermeintlich höchsten pay-off bei (D,C) realisieren zu können, sollte also zumindest berücksichtigen, daß zukünftige naturwissenschaftliche Forschungsergebnisse auch eine ganz andere Spielstruktur begründen könnten. Wenn sich in Stufe 2 herausstellt, daß die pay-offs bei einseitiger oder gar beidseitiger Nichtkooperation⁹² wesentlich niedriger liegen als ursprünglich angenommen wurde, kann eine commitment-Entscheidung in Stufe 1 suboptimal sein.

⁹¹ Die Ergebnisse aus dieser Studie werden hier nach Cline (1991) zitiert.

⁹² Wobei man sich unter Nichtkooperation hier die fehlende Bereitschaft zur Etablierung einer nationalen Politik, die der Reduktion von Treibhausgasen dient, vorstellen könnte.

2. Ein Modell

Das Spiel besteht aus zwei Stufen. In Stufe 1 steht Spieler 1 vor der Entscheidung sich an die Aktion "D" in Stufe 2 zu binden oder diese Bindung zu unterlassen. Eine Bindung sei zu Kosten α möglich. In Stufe 2 wird - abhängig von der Entscheidung von Spieler 1 in der ersten Stufe - entweder das "reduzierte" Chickenspiel, im folgenden als Teilspiel T1 bezeichnet, oder aber das "eigentliche" Chicken, entsprechend T2 genannt, gespielt. Die Aktionen der beiden Spieler werden in Weiterführung der Notation aus dem vorausgegangenen Kapitel durch $a_i(t_j)$ für die beiden Spieler $i = 1, 2$ in den beiden Stufen $j = 1, 2$ wiedergegeben. In Stufe 2 wählen die beiden Akteure ihre Aktionen simultan. Eine Strategie sei durch Φ_i gekennzeichnet. Ein teilspielperfektes Nash-Gleichgewicht N läßt sich schließlich durch das Strategieprofil $\{\Phi_i^*\}$ mit den beiden Strategien $\Phi_1^* = \{a_1(t_1), a_1(t_2)\}$ und $\Phi_2^* = \{a_2(t_2)\}$ charakterisieren. In einer ersten Version wird unterstellt, daß beide Spieler die pay-off-Struktur kennen. Das Spiel läßt sich mit Hilfe der folgenden Abbildung leicht verdeutlichen:

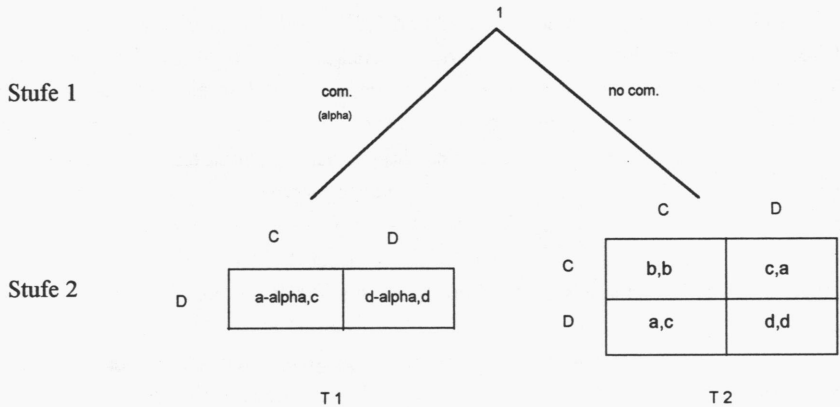


Abb.1: Chicken-Spiel mit Bindungsoption für Spieler 1

Es gilt die für ein Chickenspiel typische Konstellation, d.h. hier:

(1) $a > b > c > d$

Die extensive Form dieses Spiels kann ebenfalls in traditioneller Weise einfach dargestellt werden:

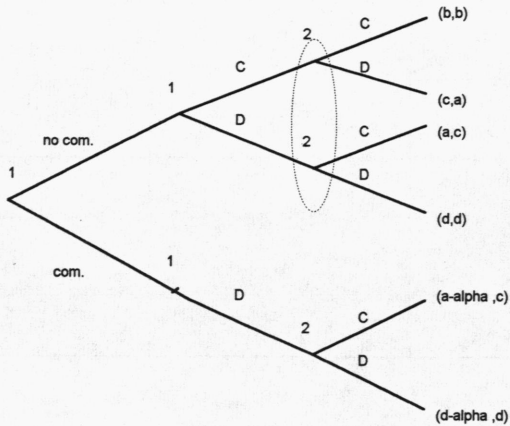


Abb.2: extensive Form des Chicken - Spiels

In diesem Modellrahmen ergeben sich vier mögliche teilspielperfekte Nash-Gleichgewichte. Wenn Spieler 1 in der ersten Runde "no com." wählt, wird in der zweiten Stufe mit T2 ein "Standard"-Chicken gespielt. In diesem Fall existieren zwei Gleichgewichte N1 und N2 in reinen Strategien sowie ein Gleichgewicht N3 in gemischten Strategien mit den pay-offs $(u_1, u_2)_{N1}$, $(u_1, u_2)_{N2}$ und $(u_1, u_2)_{N3}$.

N1	N2	N3 ⁹³
$\{\Phi_1^*\} = \{\text{"no com."}, \text{"C"}\}$ $\{\Phi_2^*\} = \{\text{"D"}\}$	$\{\Phi_1^*\} = \{\text{"no com."}, \text{"D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$	$\{\Phi_1^*\} = \{\text{"no com."}, p^* = \frac{c-d}{a+c-b-d}\}$ $\{\Phi_2^*\} = \{q^* = \frac{c-d}{a+c-b-d}\}$
$(u_1, u_2)_{N1} =$ (c, a)	$(u_1, u_2)_{N2} =$ (a, c)	$(u_1, u_2)_{N3} =$ $\left(\frac{ac-bd}{a+c-b-d}, \frac{ac-bd}{a+c-b-d} \right)$

Wählt Spieler 1 dagegen "com." in der ersten Stufe, so wird in der zweiten Runde das "reduzierte" Chicken gespielt, es existiert dann nur ein Gleichgewicht N4 in reinen Strategien

N4
$\{\Phi_1^*\} = \{\text{"com."}, \text{"D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$
$(u_1, u_2)_{N4} = (a-\alpha, c)$

Ob eine Bindung für Spieler 1 vorteilhaft ist, hängt von der Höhe von α ab. Bei einer "commitment"-Entscheidung ergibt sich $u_1 = a - \alpha$. Entschließt sich Spieler 1 jedoch

⁹³Das Gleichgewicht in gemischten Strategien erhält man aus $Eu_1(\text{"C"}) = bq+c(1-q) = aq+d(1-q) = Eu_1(\text{"D"})$. Daraus folgt q^* . p^* erhält man aus $Eu_2(\text{"C"}) = bp+c(1-p) = ap+d(1-p)$. Setzt man q^* in $Eu_1(\text{"C"})$ oder $Eu_1(\text{"D"})$ bzw. p^* in $Eu_2(\text{"C"})$ oder $Eu_2(\text{"D"})$ ein, erhält man $(u_1, u_2)_{N3}$.

gegen eine Bindung, so resultieren im folgenden Chickenspiel drei mögliche Gleichgewichte. Ohne eine zusätzliche Annahme eines weiteren Selektionsprinzips ist für Spieler 1 in der ersten Stufe nicht klar, welches der Gleichgewichte in T2 ausgewählt wird. Deswegen wird als Erwartungsnutzen nach einer "no commitment"-Wahl der Nutzen aus N1, N2 und N3, gewichtet mit der jeweils gleichen Wahrscheinlichkeit 1/3 verwendet, worin die Anwendung von Savage' "principle of insufficient reason" zum Ausdruck kommt⁹⁴. Dieser Erwartungsnutzen ergibt sich somit als

$$\begin{aligned}
 (2) \quad Eu_1(\text{"no commitment"}) &= \frac{1}{3}c + \frac{1}{3}a + \frac{1}{3} \frac{ac - bd}{a + c - b - d} \\
 &= \frac{1}{3} \left[c + a + \frac{ac - bd}{a + c - b - d} \right]
 \end{aligned}$$

Eine Bindung ist für Spieler 1 also vorteilhaft, falls

$$(3) \quad \alpha < a - \frac{1}{3} \left[c + a + \frac{ac - bd}{a + c - b - d} \right]$$

gilt.

Bei *Unsicherheit* über die pay-offs ändert sich die Entscheidungssituation. Dazu sei im folgenden angenommen, daß die beiden Spieler mit Wahrscheinlichkeit π wie bisher davon ausgehen, daß die bisher diskutierten pay-offs zu erwarten sind, mit Wahrscheinlichkeit $1 - \pi$ dagegen damit rechnen, daß bei Nichtkooperation von einem oder gar beiden Spielern die pay-offs wesentlich geringer ausfallen, hier sogar um soviel geringer, daß in Stufe 2 nach einer "no commitment" - Entscheidung des ersten Spielers anstelle eines Chickenspiels ein Harmony-Spiel⁹⁵ resultiert. Dieser Teil der modifizierten Stufe 2 mit den neuen Teilspielen T3 und T4 läßt sich durch die nächste Abbildung veranschaulichen.

⁹⁴vgl. Savage (1972), S.63ff. Siehe dort auch die Bemerkungen zur Kritik an diesem Prinzip.

⁹⁵Diese Interessenskonstellation beinhaltet "C" als dominante Strategie für beide Spieler. (C,C) ist damit eindeutiges Nash-Gleichgewicht. Die Modellierung über ein mögliches Eintreten einer *Harmony*-Struktur soll dazu dienen, die Risiken der Selbstbindung in Stufe 1 möglichst deutlich herauszustellen.

$1 - \pi$

Abb.3: Teil der modifizierten Stufe 2 bei Unsicherheit über die pay-offs

Als Rangordnung der Parameter gelte:

(4) $a > a'$; $c > c'$; $d > d'$

Dadurch wird der Gedanke modelliert, daß die Auszahlungen wesentlich geringer ausfallen, wenn mindestens einer der beiden Spieler nicht kooperiert. Damit "C" in T4 für beide Spieler dominante Strategie ist, muß weiterhin gelten:

(5) $b > a' ; c' > d'$

Die vier möglichen Nash-Gleichgewichte incl. zugehöriger pay-offs lauten jetzt:

N1	N2	N3 ⁹⁶
$\{\Phi_1^*\} =$ $\{\text{"no com.", "C"}\}$ $\{\Phi_2^*\} = \{\text{"D"}\}$	$\{\Phi_1^*\} =$ $\{\text{"no com.", "D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$	$\{\Phi_1^*\} =$ $\{\text{"no com.", } p^* =$ $\frac{\pi(c-d) + (1-\pi)(c'-d')}{\pi(a+c-d) - b + (1-\pi)(a'+c'-d')}\}$ $\{\Phi_2^*\} =$ $\{q^* =$ $\frac{\pi(c-d) + (1-\pi)(c'-d')}{\pi(a+c-d) - b + (1-\pi)(a'+c'-d')}\}$
$(u_1, u_2)_{N1} =$ $(c\pi + c'(1-\pi), a\pi + a'(1-\pi))$	$(u_1, u_2)_{N2} =$ $(a\pi + a'(1-\pi), c\pi + c'(1-\pi))$	$(u_1, u_2)_{N3} =$ $(bq^* + [c\pi + c'(1-\pi)](1-q^*),$ $bq^* + [c\pi + c'(1-\pi)](1-q^*))$

N4
$\{\Phi_1^*\} = \{\text{"com.", "D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$
$(u_1, u_2)_{N4} =$ $(a\pi + a'(1-\pi) - \alpha, c\pi + c'(1-\pi))$

⁹⁶Das Gleichgewicht in gemischten Strategien erhält man aus $Eu_1(\text{"C"}) = bq + [c\pi + c'(1-\pi)](1-q) = [a\pi + a'(1-\pi)]q + [d\pi + d'(1-\pi)](1-q) = Eu_1(\text{"D"})$. Daraus folgt q^* . p^* erhält man aus $Eu_2(\text{"C"}) = bp + [c\pi + c'(1-\pi)](1-p) = [a\pi + a'(1-\pi)]p + [d\pi + d'(1-\pi)](1-p) = Eu_2(\text{"D"})$. Bei symmetrischen pay-offs ergibt sich: $p^* = q^*$.

Die Frage nach der Selbstbindung in der ersten Stufe besteht für Spieler 1 jetzt aus einem Vergleich seiner beiden Erwartungsnutzen

$$(6) \quad Eu_1(\text{"commitment"}) = a\pi + a'(1-\pi) - \alpha$$

und

$$(7) \quad Eu_1(\text{"no commitment"}) =$$

$$\frac{1}{3} [c\pi + c'(1-\pi)] + \frac{1}{3} [a\pi + a'(1-\pi)] + \frac{1}{3} \left\{ bq^* + [(c\pi + c'(1-\pi)) (1-q^*)] \right\}$$

Bei Gleichung (7) wurde wieder Savage' "principle of insufficient reason" berücksichtigt, indem die Wahrscheinlichkeiten für die drei möglichen Chicken-Gleichgewichte identisch mit jeweils 1/3 angenommen wurden. u_1 im Gleichgewicht in gemischten Strategien ergibt sich durch Berücksichtigung von q^* in $Eu_1(\text{"C"})$ ⁹⁷. In dieser Konstellation ist eine Selbstbindung somit vorteilhaft, wenn (6) > (7) gilt.

⁹⁷Da die Werte bei Verwendung allgemeiner Parameter hier etwas unübersichtlich werden, wird im Anhang ein einfaches Zahlenbeispiel vorgestellt.

3. Fairness

Wieder beginne ich mit der Motivation des gewählten Modells, diesmal durch ein Beispiel aus der experimentellen Literatur (1.). Es folgt das Modell (2.), zuerst in einer vereinfachten Version bei Sicherheit über die Höhe des Fairnessparameters (2.1), anschließend in einer erweiterten Variante unter Berücksichtigung von Unsicherheit (2.2).

1. Das Ultimatum-Spiel

Das Risiko einer Bindung läßt sich auch vor einem anderen Hintergrund demonstrieren. Die Entscheidungen von Menschen werden in vielen Situationen durch Fairnessüberlegungen beeinflusst. Das Standardbeispiel für solches Verhalten findet sich in der experimentellen Überprüfung des Ultimatum-Spiels⁹⁸. Bei diesem Spiel macht der erste Spieler einen Vorschlag über die Aufteilung einer zur Verfügung stehenden Summe von DM x , den die zweite Spielerin akzeptieren kann oder nicht. Im Falle der Ablehnung enden beide Spieler ohne irgendeine Auszahlung. Die beste Offerte für Spieler 1 im Sinne traditioneller Rationalitätsannahmen wäre natürlich Spielerin 2 einen Pfennig anzubieten und den Rest für sich selbst zu beanspruchen, was Spielerin 2 anschließend akzeptieren würde. Die Ergebnisse unterschiedlicher Experimente zeigten jedoch einen deutlich höheren Anteil an x , der der Mitspielerin angeboten wurde. In einer der ersten Versuchsreihen stellten Güth, Schmittberger und Schwarze (1982) ein durchschnittliches Angebot von 37% der gesamten zur Verfügung stehenden Summe fest⁹⁹. Die Höhe des Anteils, die als fair erachtet wird, differiert je nach den Rahmenbedingungen der Entscheidungssituation¹⁰⁰, die bloße Existenz von Fairnessmotiven in den Präferenzen jedoch ist schwer widerlegbar.

Wenn nun die Bindung an "D" vom Mitspieler als unfair erachtet wird, kann sich der Anreiz für eine solche Entscheidung deutlich verändern. Bei Unsicherheit über die

⁹⁸ Einen Überblick über diese Ergebnisse liefert Richard Thaler (1992) in Kapitel 3 seines insgesamt höchst lesenswerten Buchs "The Winner's Curse", in dem er verschiedene Anomalien präsentiert.

⁹⁹ siehe auch Ochs/Roth (1989) und die dort zitierte experimentelle Literatur.

¹⁰⁰ vgl. dazu (wiederum) den Aufsatz von Yaari und Bar-Hillel (1984).

Stärke der Fairnessempfindungen ergibt sich - anders motiviert - wieder ein Argument für das Risiko eines self-commitments.

2. Ein Modell

Die Entscheidung zur Selbstbindung in Stufe 1 bedeutet, daß ein Spieler beidseitiger Kooperation keine Chance gibt. Ist eine solche Entscheidung immer noch rational, wenn sie vom Gegenspieler als unfair betrachtet wird? Berücksichtigt man zusätzlich zum bisherigen Modellrahmen eine mögliche Abhängigkeit menschlicher Entscheidungen von Fairnessüberlegungen, so läßt sich ein differenziertes Ergebnis diskutieren. Als Grundkonstellation wird dabei wieder ein zweistufiges Spiel gewählt. Spieler 1 steht in Stufe 1 vor der Entscheidung, sich an "D" zu binden oder nicht. In der zweiten Runde wird dann bei simultanen Offerten *unter Berücksichtigung von Fairnessmotiven* das "reduzierte" Chicken T1' (bei commitment-Entscheidung von Spieler 1) oder das Standard-Chickenspiel T2' (bei no-commitment-Entscheidung) gespielt¹⁰¹. In einer ersten Präsentation wird unterstellt, daß Spieler 1 die Stärke der Fairnessmotive von Spieler 2 kennt, in einer zweiten Version hingegen wird davon ausgegangen, daß sich Spieler 1 unsicher über die Intensität der Fairnessempfindungen beim anderen Spieler ist.

2.1 Fairness bei Sicherheit über die Höhe des Fairnessparameters

Die zentrale Frage lautet sicher, in welcher Weise sich Fairness modellieren läßt. Matthew Rabin charakterisiert Fairnessmotive vor dem Hintergrund der experimentellen Literatur der letzten 15 Jahre durch drei stilisierte Fakten:

- (A) People are willing to sacrifice their own material well-being to help those who are being kind.
- (B) People are willing to sacrifice their own material well-being to punish those who are being unkind.
- (C) Both motivations (A) and (B) have a greater effect on behavior as the material cost of sacrificing becomes smaller. "¹⁰²

¹⁰¹ siehe dazu später Abb.4. Diese Abbildung entspricht Abb.1 incl. der Erweiterung durch Fairness.

¹⁰² siehe Rabin (1993), S. 1282.

Ohne die kompliziertere Formalisierung einer Fairnessfunktion von Rabin zu benötigen, lassen sich die Konsequenzen der Berücksichtigung von Fairness vor diesem Hintergrund einfach demonstrieren¹⁰³. Dazu wird unterstellt, daß eine Bindung von Spieler 1 an "D" von Spieler 2 als "unfreundlicher Akt" im Sinne (B) empfunden wird und deswegen eine Antwort "D" als Bestrafung zusätzlichen Nutzen f_B für Spieler 2 stiftet. Der Verzicht auf eine mögliche Selbstbindung in der ersten Stufe hingegen wird als "kind"¹⁰⁴ im Sinne (A) interpretiert. Die Wahl von Kooperation liefert Spieler 2, wenn er sich in diesem Teilspiel befindet, somit ebenfalls einen zusätzlichen Nutzen in Höhe von f_A . Dieser Wert sei nicht abhängig davon, ob Spieler 1 seinerseits in der zweiten Runde "C" oder "D" wählt. Durch die Berücksichtigung von f wird lediglich die Aktion von Spieler 1 in der ersten Stufe von Spieler 2 bewertet. Der dritte stilisierte Fakt (C) läßt sich durch die Abhängigkeit des Fairnessparameters f von der pay-off-Differenz zwischen c und d , bzw. zwischen a und b modellieren: $f_B(|c - d|)$ und $f_A(|b - a|)$ mit den Ableitungen $df_B / d|c - d| < 0$, $df_A / d|b - a| < 0$. Je geringer die Differenz zwischen c und d ist, desto geringer ist der "Verlust" durch eigene Nichtkooperation im "reduzierten" Chicken-Spiel. Akzeptiert man die Verhaltenshypothese (C) von Rabin wird die Fairnessmotivation f mit geringeren Kosten immer stärker oder anders ausgedrückt: man unterstellt $df_B / d|c - d| < 0$. Entsprechend ist der "Verlust" durch Kooperation als Antwort auf gegnerische Kooperation umso geringer, je geringer die Differenz zwischen a und b ist.

Diese Fairnessmodellierung wird in der nächsten Abbildung verdeutlicht:

¹⁰³ Rabin (1993) diskutiert in seinem Aufsatz eine "sophisticated version" von Fairness. Darin verwendet er den Gedanken, daß die Bewertung einer Aktion des Gegenspielers als fair oder unfair davon abhängt, welche Vermutungen man selbst über die Vermutungen des anderen besitzt. Im klassischen Battle-of-Sexes-Beispiel "Oper/Boxen" bewerte ich die Aktion "Oper" meines Gegenspielers als fair, wenn ich vermute, daß der andere diese Aktion im Glauben gewählt hat, daß ich selbst auch Oper präferiere. Wenn ich jedoch die gegenläufige Vermutung hege, daß mein Gegenspieler im Glauben handelt, ich selbst würde Boxen bevorzugen, bewerte ich seine Wahl in die Oper zu gehen als unfair. Von einer Modellierung dieser Erkenntnis wird hier jedoch abgesehen.

¹⁰⁴ "Being Kind" bedeutet nicht nur "freundliches" Verhalten, sondern bezeichnet die Fähigkeit "sich in die Schuhe anderer Menschen zu stellen". Dadurch wird in der genaueren Übersetzung selbst ein Fairnessprinzip deutlich. Für diese Begriffsklärung möchte ich mich bei Clive Bell bedanken.

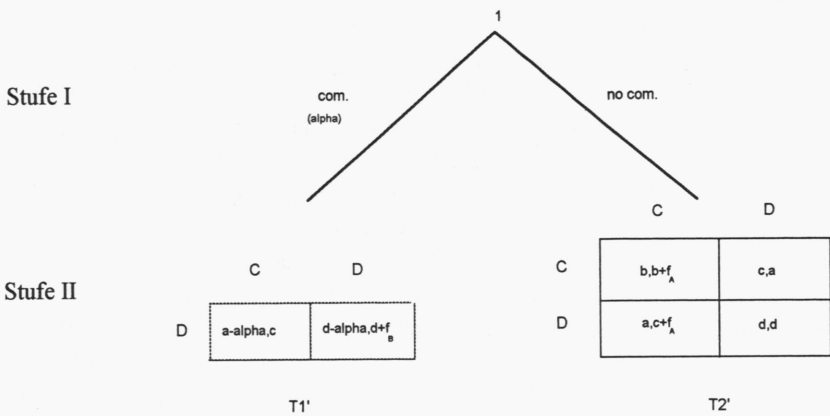


Abb.4 : commitment in einem Chickenspiel mit Fairness

Wenn Spieler 1 weiß, wie stark die Fairnessmotivation bei Spieler 2 ausgeprägt ist, ergeben sich die Gleichgewichte in Abhängigkeit von f wie folgt:

$$\text{Fall 1: } f_B < c - d \wedge f_A < a - b$$

Die erste Ungleichung $f_B < c - d$ bedeutet: in diesem Fall wird eine "commitment"-Entscheidung von Spieler 1 durch Spieler 2 nicht so "unfair" im Sinne von Motiv (B) eingestuft, daß im "reduzierten" Chicken T1' als Antwort auf $a_1(t_2) = "D"$ auf $a_2(t_2) = "D"$ gewechselt würde.

$f_A < a - b$ hat zur Folge, daß in T2' weiterhin für Spieler 2 gilt: "C" ist beste Antwort auf "D" und "D" ist beste Antwort auf "C". Hier ist die Wirkung eines Verzichts auf eine mögliche Bindung (Motiv A) nicht so stark, als daß Spieler 2 "C" als dominante Strategie hätte.

Die Gleichgewichte lassen sich analog zum vorherigen Abschnitt beschreiben. Wenn Spieler 1 in der ersten Runde "no commitment" wählt, wird in der zweiten Runde das

"Standard"-Chicken mit Fairness T2' gespielt. Die drei möglichen Gleichgewichte mit pay-offs lauten:

N1	N2	N3 ¹⁰⁵
$\{\Phi_1^*\} = \{ \text{"no com."}, "C" \}$ $\{\Phi_2^*\} = \{ "D" \}$	$\{\Phi_1^*\} = \{ \text{"no com."}, "D" \}$ $\{\Phi_2^*\} = \{ "C" \}$	$\{\Phi_1^*\} = \{ \text{"no com."}, p^* = \frac{c-d+f_A}{a+c-b-d} \}$ $\{\Phi_2^*\} = \left\{ q^* = \frac{c-d}{a+c-b-d} \right\}$
$(u_1, u_2)_{N1} = (c, a)$	$(u_1, u_2)_{N2} = (a, c+f_A)$	$(u_1, u_2)_{N3} = \left(\frac{a-bd}{a+c-b-d}, \frac{a-bd}{a+c-b-d} + \frac{f_A(a-d)}{a+c-b-d} \right)$

Im Vergleich zum Modell ohne Fairness ändert sich zum einen bei N2 der pay-off u_2 sowie zum anderen bei N3 p^* und damit auch u_2 .

Nach der Wahl von $a_1(t_1) = \text{"com."}$ in der ersten Stufe wird in der zweiten Runde das "reduzierte" Chicken T1' gespielt. Als Gleichgewicht resultiert in diesem Fall N4:

¹⁰⁵Das Gleichgewicht in gemischten Strategien erhält man aus $Eu_1("C") = bq+c(1-q) = aq+d(1-q) = Eu_1("D")$. Daraus folgt q^* . p^* erhält man aus $Eu_2("C") = (b+f_A)p+(c+f_A)(1-p) = ap+d(1-p)$. Setzt man q^* in $Eu_1("C")$ oder $Eu_1("D")$ bzw. p^* in $Eu_2("C")$ oder $Eu_2("D")$ ein, erhält man $(u_1, u_2)_{N3}$.

N4
$\{\Phi_1^*\} = \{\text{"com."}, \text{"D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$
$(u_1, u_2)_{N4} = (a - \alpha, c)$

Die Vorteilhaftigkeit einer Bindung wird über (2) wieder durch (3) beschrieben. Die Einführung eines Fairnessparameters f allein ändert somit noch nichts an der Entscheidungssituation für Spieler 1, falls f hinreichend gering ist und Spieler 1 die Größenordnung von f kennt. Das Ergebnis aus dem Modell ohne Fairness findet somit auch hier seine Anwendung, allerdings spezifiziert für einen bestimmten Bereich der Fairnessgefühle. In einer Situation, in der die Entscheidung von Spieler 2 jedoch sehr stark von Fairnessmotiven geprägt ist, präsentieren sich die Überlegungen des ersten Spielers natürlich auch in einem anderen Gewand.

$$\text{Fall 2: } f_B > c - d \wedge f_A > a - b$$

Bei hinreichend großer Fairnessmotivation im Sinne von Rabins Fakt (B), d.h. für $f_B > c - d$, läßt sich Bindung in Stufe 1 überhaupt nicht mehr begründen. Wenn Spieler 1 weiß, daß Spieler 2 so starke Fairnessmotive besitzt, daß er als Reaktion auf eine Bindungsentscheidung von Spieler 1 in der ersten Stufe in Stufe 2 "D" wählen wird, ergibt sich $u_1 = d - \alpha$. Weniger als d kann Spieler 1 bei Wahl von "no commitment" jedoch nicht bekommen. Als mögliche Gleichgewichte in T2' verbleiben N2 und N3¹⁰⁶.

$$\text{Fall 3: } f_B > c - d \wedge f_A < a - b$$

Spieler 1 wird gemäß obiger Überlegungen wieder "no com." in der ersten Stufe wählen. In T2' existieren für $f_A < a - b$ wieder die drei Gleichgewichte N1, N2 und N3.

¹⁰⁶ Bei $f_A > a - b$ ist "C" dominante Strategie für Spieler 2 im Teilspiel T2', da gleichzeitig $c > 0 > d - f_A$ gilt. Deshalb scheidet N1 als Gleichgewicht aus.

Welches der Gleichgewichte ausgewählt wird, ist ohne zusätzliche Annahmen nicht begründbar. Jeder der drei Nutzenwerte $(u_1)_{N1-N3}$ ist jedoch strikt größer als $d - \alpha$.

$$\text{Fall 4: } f_B < c - d \wedge f_A > a - b$$

Die Argumentation verläuft hier anhand von T2': für $f_A > a - b$ ist die Fairnessmotivation (A) bei Spieler 2 so stark ausgeprägt, daß eine "no commitment" - Entscheidung des ersten Spielers durch die Wahl von "C" honoriert wird. Wieder verbleiben - wie in Fall 2 - zwei mögliche Gleichgewichte: N2 und N3. Bei einer "commitment"-Entscheidung resultiert als Gleichgewicht N4. Eine Bindung wird vom ersten Spieler präferiert, falls

$$(8) \quad a - \alpha > \frac{1}{2} \left[a + \frac{ac - bd}{a + c - b - d} \right]$$

gilt.

2.2 [Fairness bei Unsicherheit über die Höhe des Fairnessparameters]

Im nächsten Schritt wird das Risiko einer Selbstbindung in Stufe 1 bei Unsicherheit bzgl. der Höhe des Fairnessparameters f_B diskutiert. Spieler 1 weiß, daß im Falle einer freiwilligen Handlungsbeschränkung diese Aktion von Spieler 2 als unfair erachtet wird, ist sich aber über die Höhe von f_B unsicher. Mit π geht er davon aus, daß $f_B > c - d$ gilt, mit $1 - \pi$ dagegen vermutet er, daß $f_B < c - d$ gilt und somit Spieler 2 im "reduzierten" Chickenspiel T1' immer noch "C" wählt. Da hier gerade das Risiko einer Selbstbindung thematisiert werden soll, wird zur Vereinfachung angenommen, daß die Existenz und Größenordnung von f_A beiden Spielern bekannt ist. Im besonderen soll gelten: $f_A < a - b$. Die Präferenzen von Spieler 2 in T2' sind also "chicken-like": "C" ist beste Antwort auf "D" und "D" ist beste Antwort auf "C". Würde man die umgekehrte Relation $f_A > a - b$ unterstellen, wäre das Problem wieder trivial (siehe die obigen Bemerkungen zu Fall 4).

Der Vergleich der beiden hier relevanten Erwartungswerte lässt sich jetzt sehr schnell verdeutlichen:

$$(9) \quad \text{Eu}_1(\text{"commitment"}) = \pi(d-\alpha) + (1-\pi)(a-\alpha) = \pi d + (1-\pi)a - \alpha$$

$$(2) \quad \text{Eu}_1(\text{"no commitment"}) = \frac{1}{3} \left[c + a + \frac{ac - bd}{a + c - b - d} \right]$$

Im Unterschied zu (3) ist eine Bindung nun vorteilhaft, falls

$$(10) \quad \alpha < \pi d + (1-\pi)a - \frac{1}{3} \left[c + a + \frac{ac - bd}{a + c - b - d} \right]$$

gilt. Bei gegebenen Kosten der Bindung α sinkt der Anreiz sich zu binden mit steigender Wahrscheinlichkeit π . Je eher Spieler 1 damit rechnet, durch seine Bindungsentcheidung seinen Gegenspieler so stark zu "motivieren", daß er seinerseits in T2' als Antwort auf "D" ebenfalls "D" spielen wird, desto eher wird natürlich Spieler 1 davon absehen, sich in der ersten Stufe zu binden.

Wir sind am Ende der Abhandlung angelangt.

Wir begannen, die Interessenslagen zwischen Menschen nach ihrem Konfliktgehalt zu unterscheiden. Die Etablierung eines marktwirtschaftlichen Systems allein ist nicht in der Lage, die angesprochenen Probleme der Kooperation und der Koordination zu lösen. Dies gilt nicht in einer friktionslosen (Arrow-Debreu)-Welt, da dort diese Probleme nicht existieren. Sie existieren jedoch sehr wohl, wenn Externalitäten vorliegen. Öffentliche Güter, zunehmende Skalenerträge und asymmetrische Informationen begründen soziale Dilemmata. Die Suche nach Auswegen führte uns zur Frage der Glaubwürdigkeit angekündigter Handlungen. Am Beispiel der axiomatischen und der strategischen Verhandlungstheorie näherten wir uns dem - aus der Glaubwürdigkeitsfrage entstehenden - Thema der Bindung.

Die Spieltheorie diente als methodisches Instrumentarium der beiden "Versuche".

Der erste "Versuch" griff das Problem der Kooperation auf.

Kooperatives Agieren einer Spielerin entstand aufgrund ihrer Vermutungen über die Konsequenzen ihres Handelns bei ihrer Gegenspielerin. Diese Konsequenzen bezogen sich auf deren Erwartungen bezüglich ihrer Präferenzen. Das Ergebnis war abhängig von den Wahrscheinlichkeiten $(p, 1-p)$ und $(\alpha, 1-\alpha)$ sowie den Nutzenwerten beider Spielerinnen. $(p, 1-p)$ war die Wahrscheinlichkeitsverteilung der zweiten Spielerin über die möglichen Typen der ersten Spielerin, $(\alpha, 1-\alpha)$ bezog sich auf den von Spielerin 2 vermuteten trade-off der ersten Spielerin zwischen einem niedrigeren Nutzen heute (durch kooperatives Handeln) und einem höheren Nutzen morgen.

Ich diskutierte verschiedene Variationen des Grundmodells. Der entscheidende Grund für die Bereitschaft zur Kooperation bei einer Spielerin mit Präferenzen, die sich durch die dominante Strategie der Nichtkooperation im stage-game charakterisieren lassen, lag in der Verbindung von eigenem Handeln und gegnerischer Erwartungsbildung.

Der zweite "Versuch" stellte sich der Frage der Koordination.

Dabei argumentierte ich zuerst in einem Modell unvollkommener Information bei einer "Chicken"-Konstellation. Die internationale Umweltpolitik lieferte mir das

Anwendungsbeispiel: die Bindung eines Staates an Nichtkooperation heute (durch eine entsprechende nationale Energiepolitik) in Unkenntnis der Nutzenwerte morgen. (ausgedrückt durch die klimatischen Veränderungen, verursacht durch die Emission von CO_2). Da das Modell äußerst einfach gehalten ist, war das Ergebnis nicht sehr überraschend. Die Vorteilhaftigkeit einer Bindung ist abhängig von den Präferenzen, von $(\pi, 1-\pi)$ und von α . Dabei bezeichnete α in diesem Modell die Bindungskosten und $(\pi, 1-\pi)$ die Wahrscheinlichkeitsverteilung über die beiden möglichen pay-off-Zustände in der zweiten Stufe.

Das zweite Modell legte asymmetrische Informationen zugrunde. Ein Spieler besaß wieder die Option der Bindung in einer ersten Stufe, wußte auch, daß er durch seine Entscheidung die Fairnessgefühle des anderen Spielers beeinflusste, wußte jedoch nicht, in welcher Intensität. Wieder folgte, daß eine Bindungsentscheidung in Abhängigkeit der Präferenzen sowie π und α getroffen wird.

Die beiden "Versuche" beinhalten nicht sämtliche Erscheinungsformen sozialer Dilemmata, sie setzen Akzente. Sie wurden gewählt, weil mit ihrer Hilfe versucht wurde, die Bedeutung von Erwartungsbildung, Glaubwürdigkeit und Bindung aufzuzeigen.

Ich schließe, wie ich begann:

Es scheint, daß die Einheit, nach der wir streben, sich ohne unser Wissen bildet und innerhalb der Grenzen sich kundgibt, die wir unserem Werk auferlegen. ... Wieviel gesünder und heilsamer ist es, nach der Realität eines Begrenzten zu trachten als nach der Unendlichkeit des Ungeeinten!

Meinen Sie nun, daß ich das Loblied der Monotonie anstimme? ... In Wahrheit ist keine Verwechslung möglich zwischen der Monotonie, die aus dem Mangel an Mannigfaltigkeit entsteht, und der Einheit, die eine Harmonie von Mannigfaltigkeiten ist - ein Maß des Vielfältigen."

Igor Strawinsky (1983), S.255

Anhang

Ein einfaches Zahlenbeispiel zu Abschnitt 2 aus dem "Versuch über die Idee der Bindung"

a, b, c, d sowie a', b', c', d' seien durch folgende Werte gegeben:

a	b = b'	c	d	a'	c'	d'
8	6	4	2	2	1	0

Mit Wahrscheinlichkeit π wird dann entweder das "reduzierte" Chickenspiel (T1) oder das "eigentliche" Chickenspiel T2 gespielt, je nachdem, ob Spieler 1 in der ersten Runde "commitment" oder "no commitment" wählt.

	C	D
D	$8-\alpha, 4$	$2-\alpha, 2$

T 1

	C	D
C	6,6	4,8
D	8,4	2,2

T 2

Mit Wahrscheinlichkeit $1 - \pi$ wird dagegen die modifizierte Stufe 2 gespielt.

	C	D
D	$2-\alpha, 1$	$-\alpha, 0$

T 3

	C	D
C	6,6	1,2
D	2,1	0,0

T 4

Bei **Sicherheit** ($\pi = 1$) und Wahl von $a_1(t_1) = \text{"no com."}$ ergeben sich als mögliche Gleichgewichte N1, N2 und N3 mit den zugehörigen pay-offs

N1	N2	N3
$\{\Phi_1^*\} = \{\text{"no com."}, \text{"C"}\}$ $\{\Phi_2^*\} = \{\text{"D"}\}$	$\{\Phi_1^*\} = \{\text{"no com."}, \text{"D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$	$\{\Phi_1^*\} = \{\text{"no com."}, p^*=1/2\}$ $\{\Phi_2^*\} = \{q^* = 1/2\}$
$(u_1, u_2)_{N1} = (4,8)$	$(u_1, u_2)_{N2} = (8,4)$	$(u_1, u_2)_{N3} = (5,5)$

Bei $a_1(t_1) = \text{"com."}$ wird das Gleichgewicht N4 in reinen Strategien beschrieben durch:

N4
$\{\Phi_1^*\} = \{\text{"com."}, \text{"D"}\}$ $\{\Phi_2^*\} = \{\text{"C"}\}$
$(u_1, u_2)_{N4} = (8-\alpha,4)$

Gleichung (2) wird damit zu

$$(2') \qquad Eu_1(\text{"no commitment"}) \quad = \quad \frac{1}{3}4 + \frac{1}{3}8 + \frac{1}{3}5 = 5,\bar{6}$$

Eine Bindung ist also vorteilhaft, falls

$$(3') \qquad \alpha < 2,\bar{3}$$

gilt.

Bei **Unsicherheit** ändern sich die Gleichgewichte N1 - N4 wie folgt:

N1	N2	N3
$\{\Phi_1^*\} =$ $\{ \text{"no com."}, "C" \}$ $\{\Phi_2^*\} = \{ "D" \}$	$\{\Phi_1^*\} =$ $\{ \text{"no com."}, "D" \}$ $\{\Phi_2^*\} = \{ "C" \}$	$\{\Phi_1^*\} =$ $\{ \text{"no com."}, p^* = \frac{\pi + 1}{7\pi - 3} \}$ $\{\Phi_2^*\} = \{ q^* = \frac{\pi + 1}{7\pi - 3} \}$
$(u_1, u_2)_{N1} =$ $(3\pi + 1, 6\pi + 2)$	$(u_1, u_2)_{N2} =$ $(6\pi + 2, 3\pi + 1)$	$(u_1, u_2)_{N3} =$ $(\frac{2 + 18\pi^2}{7\pi - 3}, \frac{2 + 18\pi^2}{7\pi - 3})$

Durch die Berücksichtigung von (p^*, q^*) in $Eu_1("C") = Eu_1("D")$ bzw. in $Eu_2("C") = Eu_2("D")$ ergibt sich $(u_1, u_2)_{N3}$. N4 läßt sich beschreiben durch:

N4
$\{\Phi_1^*\} = \{ \text{"com."}, "D" \}$ $\{\Phi_2^*\} = \{ "C" \}$
$(u_1, u_2)_{N4} =$ $(6\pi + 2 - \alpha, 3\pi + 1)$

Die Vorteilhaftigkeit einer Bindung wird anhand des Vergleichs von (6') und (7') verdeutlicht.

$$(6') \quad \text{Eu}_1(\text{"commitment"}) = 6\pi + 2 - \alpha$$

und

$$\begin{aligned} (7') \quad \text{Eu}_1(\text{"no commitment"}) &= \frac{1}{3}(3\pi + 1) + \frac{1}{3}(6\pi + 2) + \frac{1}{3} \left[\frac{2 + 18\pi^2}{7\pi - 3} \right] \\ &= \frac{81\pi^2 - 6\pi - 7}{21\pi - 9} \end{aligned}$$

Die Wahl einer Bindungsstrategie lohnt sich also, falls

$$(8') \quad \alpha < \frac{45\pi^2 - 6\pi - 11}{21\pi - 9}$$

gilt.

LITERATUR

- Acham, Karl (1984): "Über einige Rationalitätskonzeptionen in den Sozialwissenschaften", in: Schnädelbach, Herbert (Hg.): Rationalität, Suhrkamp
- Akerlof, George (1970): "The Market for Lemons: Quality Uncertainty and the Market Mechanism", in: Quarterly Journal of Economics, Vol.84, S.488-500
- Althammer, Wilhelm und Wolfgang Buchholz (1993): "Internationaler Umweltschutz als Koordinationsproblem", in: Adolf Wagner (Hg.): Dezentrale Entscheidungsfindung bei externen Effekten, Francke Verlag
- Althammer, Wilhelm und Wolfgang Buchholz (1995): "Die Bereitstellung eines öffentlichen Gutes aus spieltheoretischer Sicht: Die Grundsachverhalte", in: Ökonomie und Gesellschaft, Jahrbuch 12: Soziale Kooperation, S.92-129, Campus
- Andreoni, James (1990): "Impure altruism and donations to public goods: a theory of warm glow giving", in: Economic Journal, Vol.100, S.464-477
- Andreoni, James und John H. Miller (1993): "Rational Cooperation in the finitely repeated Prisoner's Dilemma: Experimental Evidence", in: Economic Journal, Vol.103, S. 570-585
- Antonides, Gerrit (1991): "Psychological Variables in Negotiation", in: Kyklos, Vol.44, Fasc.3, S.347-362
- Argyle, Michael (1991): Cooperation, the basis of sociability, Routledge
- Arrow, Kenneth und Frank Hahn (1971): General Competitive Analysis, Holden-Day Inc., San Francisco und Oliver & Boyd, Edinburgh
- Aumann, Robert (1985): "What is game theory trying to accomplish ?", in: Kenneth Arrow und Seppo Honkapohja (Hg.): Frontiers of Economics, Basil Blackwell

- Axelrod, Robert (1984): *The Evolution of Cooperation*, Basic Books
- Barr, Nicholas (1992): "Economic Theory and the Welfare State: A Survey and Interpretation", in: *Journal of Economic Literature*, Vol.30, No.2, S.741-803
- Batson, Daniel C. (1991): *The Altruism-Question*, Lawrence Erlbaum Associates, Publishers
- Bicchieri, Cristina (1992): "Two Kinds of Rationality", in: de Marchi, Neil (Hg.): *Post-Popperian Methodology of Economics*, Kluwer Academic Publishers
- Bicchieri, Cristina (1993): *Rationality and Coordination*, Cambridge University Press
- Binmore, Ken (1987): "Nash Bargaining Theory I", in: Binmore, Ken und Partha Dasgupta (Hg.): *The Economics of Bargaining*, Basil Blackwell
- Binmore, Ken (1990): *Essays on the Foundations of Game Theory*, Basil Blackwell
- Binmore, Ken (1992): *Fun and Games. A Text on Game Theory*, Heath and Company
- Binmore, Ken, Ariel Rubinstein und Asher Wolinsky (1986): "The Nash bargaining solution in economic modelling", in: *Rand Journal of Economics*, Vol.17, No.2, S.176-188
- Binmore, Ken, Martin Osborne und Ariel Rubinstein (1992): "Noncooperative Models of Bargaining", in: Aumann, Robert und Sergiu Hart (Hg.): *Handbook of Game Theory*, Elsevier Science Publishers
- Boyer, Robert und Andre Orlean (1992): "How do conventions evolve?", in: *Journal of Evolutionary Economics*, Vol.2, S.165-177
- Buber, Martin (1994): *Zwei Glaubenswelten*, Lambert Schneider
- Buchholz, Wolfgang und Christian Haslbeck (1991/1992): "Private Verhandlungen und staatliche Regulierung bei asymmetrischer Information", in: *Finanzarchiv*, Nr.49, S.167-180

- Cline, William R. (1991): "Scientific Basis for the Greenhouse Effect", in: *The Economic Journal*, Vol.101, S.904-919
- Coase, Ronald (1960): "The Problem of Social Cost", in: *Journal of Law and Economics*, Vol.3, S.1-44
- Cooper, Russell und Andrew John (1988): "Coordinating Coordination Failures in Keynesian Models", in: *Quarterly Journal of Economics*, Vol.103, Issue 3, S.441-463
- Cooper, Russell; Douglas V. DeJong; Robert Forsythe und Thomas W. Ross (1989): "Communication in the battle of the sexes game: some experimental results", in: *Rand Journal of Economics*, Vol.20, No.4, S.568-587
- Cooper, Russell; Douglas V. DeJong; Robert Forsythe und Thomas W. Ross (1990): "Selection Criteria in Coordination Games: Some Experimental Results", in: *American Economic Review*, Vol.80, No.1, S.218-231
- Cooper, Russell; Douglas V. DeJong; Robert Forsythe und Thomas W. Ross (1993): "Forward Induction in the Battle-of-the-Sexes Games", in: *American Economic Review*, Vol.5, S.1303-1317
- d'Aspremont, Claude und L. Gerard-Varet (1979): "Incentives and Incomplete Information", in: *Journal of Public Economics*, Vol.11, S.25-45
- Debreu, Gerard (1959): *Theory of Value*, New Haven
- Elster, Jon (1987): *Subversion der Rationalität*, Campus
- Elster, Jon (1989): *The cement of society*, Cambridge University Press
- Farrell, Joseph (1987): "Information and the Coase Theorem", in: *Journal of Economic Perspectives*, Vol.2, S.113-129
- Feyerabend, Paul (1983): *Wider den Methodenzwang*, Suhrkamp

- Frank, Robert (1987): "If homo oeconomicus could choose his own utility function, would he want one with a conscience?", in: *American Economic Review*, Vol.77, No.4, S. 593ff
- Frank, Robert (1988): *Passions within Reason*, W.W.Norton, New York
- Frank, Robert (1989): "Frames of Reference and the Quality of Live", in: *American Economic Review*, Vol.79, No.2, S.80-86
- Friedman, James (1986): *Game Theory with Applications to Economics*, Oxford University Press
- Fudenberg, Drew (1992): "Explaining cooperation and commitment in repeated games", in: Jean-Jaques Laffont (Hg.): *Advances in Economic Theory*, Sixth World Congress, Cambridge University Press
- Fudenberg, Drew und David K. Levine (1989): "Reputation and equilibrium selection in games with a patient player", in: *Econometrica*, Vol.57, No.4, S.759-778
- Fudenberg, Drew und Jean Tirole (1991): *Game Theory*, MIT Press
- Gauthier, David (1986): *Morals by Agreement*, Clarendon Press
- Gordon, Robert J. (1990): "What is New-Keynesian Economics?", in: *Journal of Economic Literature*, Vol. 28, September 1990, S. 1115-1171
- Güth, Werner, Rolf Schmittberger und Bernd Schwarze (1982): "An Experimental Analysis of Ultimatum Bargaining", in: *Journal of Economic Behavior and Organization*, Vol.3, S. 367-388
- Hargreaves Heap, Shaun (1992a): *The New Keynesian Macroeconomics*, Edward Elgar
- Hargreaves Heap, Shaun (1992b): "Rationality", in: Hargreaves Heap et al. (Hg.): *The Theory of Choice*, Blackwell
- Haslbeck, Christian (1994): *Zentrale versus dezentrale Internalisierung externer Effekte bei unvollständiger Information*, Lang Verlag

- Hinde, Robert A. und Jo Groebel (1991): *Cooperation and social behaviour*, Cambridge University Press
- Holländer, Heinz (1990): "A Social Exchange Approach to Voluntary Cooperation", in: *American Economic Review*, Vol.80, No.5, S.1157-1167
- Illing, Gerhard (1992): "Private Information as Transaction Costs: The Coase Theorem Revisited", in: *Journal of Institutional and Theoretical Economics*, Vol.148, S.558-576
- Inman, Robert P. (1987): "Markets, Governments, and the "New" Political Economy", in: Auerbach, Alan J. und Martin S. Feldstein (Hg.): *Handbook of Public Economics*, Elsevier Science Publishers
- Kahneman, David und Amos Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk", in: *Econometrica*, Vol.47, S.263-291
- Kalai, Ehud (1985): "Solutions to the Bargaining Problem", in: Hurwicz, Leonid, David Schmeidler und Hugo Sonnenschein (Hg.): *Social Goals and Social Organization*, Cambridge University Press
- Kalai, Ehud und M. Smorodinsky (1975): "Other Solutions to Nash's Bargaining Problem", in: *Econometrica*, Vol.43, S.513-518
- Kennan, John und Robert Wilson (1993): "Bargaining with Private Information", in: *Journal of Economic Literature*, Vol.31, No.1, S.45-104
- Kreps (1990a): "Corporate culture and economic theory", in: Alt, James und Kenneth Shepsle (Hg.): *Perspectives on positive political economy*, Cambridge University Press
- Kreps, David (1990b): *A Course in Microeconomic Theory*, Harvester Wheatsheaf
- Kreps, David (1990c): *Game Theory and Economic Modelling*, Clarendon Press
- Kreps, David und Robert Wilson (1982): "Sequential Equilibrium", in: *Econometrica*, Vol. 50, S.863-894

- Kreps, David, Paul Milgrom, John Roberts und Robert Wilson (1982): "Rational cooperation in the finitely repeated prisoners' dilemma", in: *Journal of Economic Theory*, Vol.27, S.245-252
- Laffont, Jean-Jaques (1988): *Fundamentals of Public Economics*, MIT Press
- Lewis, David (1969): *Convention: A Philosophical Study*, Harvard University Press
- Luce, R. Duncan und Howard Raiffa (1957): *Games and Decisions*, Wiley
- Maynard Smith, John (1982): *Evolution and the Theory of Games*, Cambridge University Press
- Mertens, Jean-Francois (1990): "Repeated Games", in: Ichiishi, Tatsuro, Abraham Neyman und Yair Tauman (Hg.): *Game Theory and Applications*, Academic Press
- Miller, Gary (1992): *Managerial Dilemmas*, Cambridge University Press
- Moulin, Hervé (1989): *Axioms of Cooperative Decision Making*, Cambridge University Press
- Myerson, Roger und Mark Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading", in: *Journal of Economic Theory*, Vol.28, S.265-281
- Nash, John (1950): "The Bargaining Problem", in: *Econometrica*, Vol.18, S.155-162
- Nash, John (1951): "Non-Cooperative Games", in: *Annals of Mathematics*, Vol.54, S. 286-295
- Nash, John (1953): "Two-Person Cooperative Games", in: *Econometrica*, Vol.21, S.128-140
- Ochs, Jack und Alvin E. Roth (1989): "An Experimental Study of Sequential Bargaining", in: *American Economic Review*, Vol.79, Vol.3, S.355-384
- Osborne, Martin und Ariel Rubinstein (1990): *Bargaining and Markets*, Academic Press

- Pearce, David (1992): "Repeated games: cooperation and rationality", in: Jean-Jaques Laffont (Hg.): *Advances in Economic Theory, Sixth World Congress*, Cambridge University Press
- Persson, Torsten und Guido Tabellini (1990): *Macroeconomic Policy, Credibility and Politics*, harwood academic publishers
- Rabin, Matthew (1993): "Incorporating Fairness into Game Theory and Economics", in: *American Economic Review*, Vol.83, No.5, S.1281-1303
- Rasmusen, Eric (1989): *Games and Information*, Basil Blackwell
- Roemer, John (1986): "The Mismatch of Bargaining Theory and Distributive Justice", in: *Ethics*, Vol.97, S.88-110
- Rubinstein, Ariel (1982): "Perfect Equilibrium in a Bargaining Model", in: *Econometrica*, Vol.50, S. 97-109
- Rubinstein, Ariel (1991): "Comments on the Interpretation of Game Theory", in: *Econometrica*, Vol.59, No.4, S.909-924
- Rubinstein, Ariel (1992): "Comments on the Interpretation of Repeated Game Theory", in: Jean-Jaques Laffont (Hg.): *Advances in Economic Theory, Sixth World Congress*, Cambridge University Press
- Sabourian, H. (1989): "Repeated Games: A Survey", in: Frank Hahn (Hg.): *The Economics of Missing Markets, Information, and Games*, Clarendon Press
- Samuelson, Paul A. (1993): "Altruism as a Problem Involving Group versus Individual Selection in Economics and Biology", in: *American Economic Review*, Vol.83, No.2, S.143-148
- Savage, Leonard J. (1972): *The Foundations of Statistics*, 2nd Edition, Dover Publications Inc.
- Schelling, Thomas (1956): "An Essay on Bargaining", in: *American Economic Review*, Vol. 46, No.3, S.281-306

- Sen, Amartya (1967): "Isolation, Assurance and the Social Rate of Discount", in: Quarterly Journal of Economics, Vol.81, S.112-124
- Silvestre, Joaquim (1993): "The Market-Power Foundations of Macroeconomic Policy", in: Journal of Economic Literature, Vol.31, No.1, S.105-141
- Simon, Herbert A. (1976): "From Substantive to Procedural Rationality", in: Spiro Latsis (Hg.): Method and Appraisal in Economics, Cambridge University Press
- Simon, Herbert A. (1978): "Rationality as Process and as Product of Thought", in: American Economic Review, Papers and Proceedings, Vol.68, S.1-16
- Simon, Herbert A. (1993): "Altruism and Economics", in: American Economic Review, Vol.83, No.2, S.156-161
- Spence, A. Michael (1973): "Job Market Signalling", in: Quarterly Journal of Economics, Vol.87, S.355-374
- Strawinsky, Igor (1983): Schriften und Gespräche I, Schott
- Sugden, Robert (1982): "On the Economics of Philanthropy", in: Economic Journal, Vol.92, S.341-350
- Sugden, Robert (1984): "Reciprocity: The supply of public goods through voluntary contributions", in: Economic Journal, Vol.94, S.772-787
- Sugden, Robert (1986): The Economics of Rights, Co-operation and Welfare, Basil Blackwell
- Sugden, Robert (1991): "Rational Choice: A Survey of Contributions from Economics and Philosophy", in: Economic Journal, Vol.101, S. 751-785
- Sutton, John (1986): "Non-Cooperative Bargaining Theory: An Introduction", in: Review of Economic Studies, Vol.53, S.709-724
- Taylor, Michael (1987): The Possibility of Cooperation, Cambridge University Press

- Thaler, Richard H. (1992): *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, The Free Press, New York
- Tirole, Jean (1988): *The Theory of Industrial Organization*, MIT Press
- Tversky, Amos und Daniel Kahneman (1991): "Loss Aversion in Riskless Choice. A Reference Dependent Model", in: *Quarterly Journal of Economics*, Vol.107, No.4, S.1039-1061
- Tversky, Amos und Daniel Kahneman (1992): "Advances in Prospect Theory. Cumulative Representation of Uncertainty", in: *Journal of Risk and Uncertainty*, Vol.5, S.297-323
- van Huyck, John B. et al. (1990): "Tacit coordination games, Strategic uncertainty, and coordination failure", in: *American Economic Review*, Vol. 80, No. 1, S.234 - 248
- Vogt, Winfried (1993): "Über die Rationalität der ökonomischen Theorie", in: *Ökonomie und Gesellschaft, Jahrbuch 10: Die ökonomische Wissenschaft und ihr Betrieb*, S.32-60, Campus
- Wärneryd, Karl (1990): *Economic Conventions, Essays in Institutional Evolution*, Dissertation an der Stockholm School of Economics
- Williamson, Oliver E. (1989): "Transaction Cost Economics", in: Schmalensee, Richard und Robert Willig (Hg.): *Handbook of Industrial Organization*, Vol.1, S.135-182, North-Holland
- Wilson, Robert (1985): "Reputation in games and markets", in: Alvin E.Roth (Hg.): *Game-theoretic models of bargaining*, Cambridge University Press
- Yaari, Menahem und Maya Bar-Hillel (1984): "On Dividing Justly", in: *Social Choice and Welfare*, Vol.1, S.1-24

Der Wiederaufbau im Osten Europas

Schriften aus der Reihe Wirtschafts- und Sozialwissenschaften

- Band 8 *Frank Schulz-Nieswandt*
Transformation, Modernisierung und Unterentwicklung
Zur Grundlegung einer Problemsichtweise
123 Seiten, 1994, Pb, DM 26.90
ISBN 3-929318-13-X
- Band 12 *Holger Kern*
Das Bankensystem der Tschechischen Republik
Rückblick - Transformation - Weiterentwicklung
195 Seiten, 1994, 32 Abb. und Tab., Pb, DM 39.90
ISBN 3-929318-21-0
- Band 13 *Frank Schulz-Nieswandt*
Zur Theorie der Transformation
Die "social welfare policy matters"-These unter Beachtung komparativer wirtschafts- und sozialgeographischer Aspekte
86 Seiten, 1994, Pb, DM 19.90
ISBN 3-929318-22-9
- Band 18 *Gordon P. Müller-Eschenbach*
Die Zukunftsperspektiven der Tschechischen Republik und der Slowakei zwei Jahre nach Spaltung der CSFR
- Eine volkswirtschaftliche Analyse -
186 Seiten, April 1995, 92 Abb. und Tab., Pb, DM 39.90
ISBN 3-929318-30-X
- Band 21 *Thomas Schmucker*
Wirtschaftlicher Wiederaufbau Osteuropas am Beispiel der Tschechischen Republik
86 Seiten, Juni 1995, Pb, DM 26.90
ISBN 3-929318-34-2
- Band 24 *Claudia Löhnig*
Systemtransformation
Aspekte der wirtschaftlichen Umstrukturierung in den Reformstaaten Osteuropas und der ehemaligen Sowjetunion
212 Seiten, Februar 1996, 44 Abb. und Tab. Pb, DM 49.90
ISBN 3-929318-38-5
- Band 25 *Petra Stemmer*
Abrüstung als sozialökonomisches Problem in den Transformationsländern
Eine Literatursauswertung über relevante Dimensionen und Effekte
214 Seiten, April 1996, 18 Abb., Pb, DM 49.90
ISBN 3-929318-39-3

eurotrans-Verlag, Postfach, 92608 Weiden